JC542

•
Pros.
10
må
10
10
12,
14
13
34
1-6
zá:
12
7 mi
.55

100

Please type a plus sign (+) inside this box -> +

PTO/SB/05 (4/98)
Approved for use through 09/30/2000. OMB 0651-0032
Patent and Trademark Office: U.S. DEPARTMENT OF COMMERCE
olilection of information unless it displays a valid OMB control number.

UTILITY PATENT APPLICATION **TRANSMITTAL** nonprovisional applications under 37 C.F.R. § 1.53(b).

Attornev Docket No. JB0800 First Inventor or Application Identifier Malcolm et al

Express Mail Label No.

Covalent Complexes of HCV NS3 Protease EL226882780US

APPLICATION ELEMENTS See MPEP chapter 600 concerning utility patent application contents.		ADDRESS TO: Box Patent Application Washington, DC 20231	
	ee Transmittal Form (e.g., PTO/SB/17) ubmit an original and a duplicate for fee processing)	Microfiche Computer Program (Appendix)	
1	ecification [Total Pages 59]	Nucleotide and/or Amino Acid Sequence Submission	
(pr	eferred arrangement set forth below)	(if applicable, all necessary) a. X Computer Readable Copy	
	Descriptive title of the Invention Cross References to Related Applications	= = = = = = = = = = = = = = = = = = = =	
	tatement Regarding Fed sponsored R & D	b. X Paper Copy (identical to computer copy)	
1	teference to Microfiche Appendix	c. X Statement verifying identity of above copies	
	lackground of the Invention Brief Summary of the Invention	ACCOMPANYING APPLICATION PARTS	
	rief Description of the Drawings (if filed)	Assignment Papers (cover sheet & document(s))	
	Petailed Description	8. 37 C.F.R.§3.73(b) Statement Power of Attorney	
- 0	Claim(s)	English Translation Document (if applicable)	
	bstract of the Disclosure	10. Information Disclosure Copies of IDS	
3. X Dr	awing(s) (35 U.S.C. 113) [Total Sheets 7]	10. Statement (IDS)/PTO-1449 Citations	
4. Oath or I	Declaration [Total Pages]	11. Preliminary Amendment	
a. [X Newly executed (original or copy)	12. X Return Receipt Postcard (MPEP 503) (Should be specifically itemized)	
b.	Copy from a prior application (37 C.F.R. § 1.60 (for continuation/divisional with Box 16 completed)	3(d)) Small Entity Charles of the second state	
-	i. DELETION OF INVENTOR(S)	13. Statement(s) Statement filed in prior application, Status still proper and desired	
	Signed statement attached deleting	Certified Copy of Priority Document(s)	
inventor(s) named in the prior application, see 37 C.F.R. §§ 1.63(d)(2) and 1.33(b).			
NOTE FOR ITEMS 1 & 13: IN ORDER TO BE ENTITLED TO DAY SMALL ENTITY			
FEES, A SMALL ENTITY STATEMENT IS REQUIRED (37 C.F.R. § 1.27), EXCEPT IF ONE FILED IN A PRIOR APPLICATION IS RELIED UPON (37 C.F.R. § 1.28).			
16. If a CONTINUING APPLICATION, check appropriate box, and supply the requisite information below and in a preliminary amendment:			
Continuation Divisional Continuation-in-part (CIP) of prior application No: Prior application information: Examiner Group / Art Unit:			
For CONTINUATION or DIVISIONAL APPS only: The entire disclosure of the prior application, from which an oath or declaration is supplied under Box 4b, is considered a part of the disclosure of the accompanying continuation or divisional application and is hereby incorporated by			
reference. The Incorporation can only be relied upon when a portion has been inadvertently omitted from the submitted application parts.			
17. CORRESPONDENCE ADDRESS			
☐ Customer Number or Bar Code Laber or ☐ Correspondence address below			
(Insert Customer No. or Attach bar code label here)			
Name	Jaye P. McLaughlin		
	Schering-Plough Corporati	Lon	
Address	Address 2000 Galloping Hill Road		
City	Kenilworth State	New Jersey 76 Code 07033	
Country	USA Telephone	Zip Code	
1 (300) 270-3030 124 (300) 270-3380			
Name (Print/Type) Jaye P. McLaughlin Registration No. (Attorney/Agent) 41,211			
Signature AMONIALISKUM) Date 11124198			

Burden Hour Statement: This tolym and sestimated to take 0.2 (haurs to complete. Time will vary depending upon the needs of the individual case. Any comments on the amount of timely out are required to complete this form should be sent to the Chief Information Officer, Patent and Trademark Office, Washington, Co. 20231. DO Net? SEND FEES OR COMMITTEE OF PORMS TO THIS ADDRESS. SEND TO: Assistant Commission for Patents. Box Patent Application, Washington, DC 20231.

SINGLE-CHAIN RECOMBINANT COMPLEXES OF HEPATITIS C VIRUS NS3 PROTEASE AND NS4A COFACTOR PEPTIDE

This filing is a conversion of Provisional U.S. Patent Applications USSN 60/067,315, filed November 28, 1997 and USSN 60/094,331, filed July 28, 1998, each of which is incorporated herein by reference, to a U.S. Utility Patent Application.

BACKGROUND OF THE INVENTION

Hepatitis C virus (HCV) is considered to be the major etiological agent of non-A non-B (NANB) hepatitis, chronic liver disease, and hepatocellular carcinoma (HCC) around the world, with an estimated human seroprevalence of 1% globally. [Alter et al., 1994, Castroenterol. Clin. North Am. 23:437-455; Behrens et al., 1996, EMBO J. 15:12-22]. Four million individuals may be infected in the United States. The viral infection accounts for greater than 90% of transfusion-associated hepatitis in the U.S. and it is the predominant form of hepatitis in adults over 40 years of age. Almost all of the infections result in chronic hepatitis and nearly 20% of those infected develop liver cirrhosis.

The virus particle has not been identified due to the lack of an efficient *ex vivo* replication system and the extremely low amount of HCV particles in infected liver tissues or blood. However, molecular cloning of the viral genome has been accomplished by isolating the messenger RNA (mRNA) from the serum of infected chimpanzees and preparing cDNA using recombinant methodologies. [Grakoui A. *et al.*, 1993, J. Virol. 67: 1385-1395]. It is now known that HCV contains a positive strand RNA genome comprising approximately 9400 nucleotides, organization of which is similar to that of flaviviruses and pestiviruses. The genome of HCV, a (+)-stranded RNA molecule of ~9.4 kb, encodes a single large polyprotein of about 3000 amino acids which undergoes proteolysis to form mature viral proteins in infected cells.

Cell-free translation of the viral polyprotein and cell culture expression studies have established that the HCV polyprotein is

10

15

20

25

5

30

35

25

30

35

processed by cellular and viral proteases to produce the putative structural and nonstructural (NS) proteins. At least ten mature viral proteins are produced from the polyprotein by specific proteolysis. The order and nomenclature of the cleavage products are as follows: NH2-C-E1-E2-p7-NS2-NS4A-NS3-NS4B-NS5A-NS5B-COOH (Fig. 1) [Grakoui et al., 1993, J. Virol. 67:1385-95; Hijikata et al., 1991, PNAS 88:5547-51; Lin et al., 1994, J. Virol. 68:5063-73]. The three amino-terminal putative structural proteins, C (capsid), E1, and E2 (two envelope glycoproteins), are believed to be cleaved by a host signal peptidase of the endoplasmic 10 reticulum (ER). The host enzyme is also responsible for generating the amino terminus of NS2. The proteolytic processing of the nonstructural proteins are carried out by the viral proteases: NS2-3 and NS3, contained within the viral polyprotein. The NS2-3 protease catalyzes the cleavage between NS2 and NS3. It is a metalloprotease and requires both NS2 and 15 the protease domain of NS3.

The NS3 protease catalyzes the rest of the cleavages in the nonstructural part of the polyprotein. The NS3 protein contains 631 amino acid residues and is comprised of two enzymatic activities: the protease domain contained within amino acid residues 1-181 and a helicase ATPase domain contained within the rest of the protein Kim et al., 1995, Biochem Biophys Res. Comm., 215:160-166. It is not known if the 70 kD NS3 protein is cleaved further in infected cells to separate the protease domain from the helicase domain, although no cleavage has been observed in cell culture expression studies.

The NS3 protease is a member of the serine class of enzymes. It uses a His, Asp, Ser catalytic triad. Mutation of the Ser residue abolishes cleavage of NS3/4A, NS4A/4B, NS4B/5A, and NS5A/5B substrates. The cleavage between NS3 and NS4A is intramolecular, whereas the cleavages at the NS 4A/4B, 4B/5A, 5A/5B sites occur in $\it trans$.

Experiments using transient expression of various forms of HCV NS polyproteins in mammalian cells have established that the NS3 serine protease is necessary but not sufficient for efficient processing of all of these cleavages. Like the flaviviruses, the HCV NS3 protease also requires a cofactor to catalyze some of these cleavage reactions. Efficient proteolytic processing at NS3/4A, NS4A/4B, NS4B/5A, and NS5A/5B

15

20

25

30

35

sites within the non-structural domain of hepatitis C virus requires a heterodimeric complex of the NS3 serine protease and the NS4A protein. [Bartenschlager et al. 1995, J. Virol. 67:3835-3844; Failla et al., 1994, J. Virol. 68:3753-3760]. A 13-amino acid synthetic NS4A peptide, corresponding to the central hydrophobic domain of NS4A protein, spanning residues 21-33 has been shown to be sufficient for activation of NS3 protease [Butkiewicz et al., 1996, Virology, 225: 328-338]. A smaller domain (amino acid residues 22-30) of NS4A has been shown to be sufficient for activation of the protease [Lin et al., 1995, J. Virol 69:4377-80].

The recently published three dimensional structure of the NS3 protease [Kim et al, 1996, Cell 87:343-355; Love et al, 1996, Cell 87:331-342] revealed that the N-terminal 37 residues of NS3 adopt a β (residues 6-9)- α (residues 14-22)- β (residues 33-37) structure upon binding of a synthetic peptide corresponding to the central hydrophobic domain spanning residues 21-32 of NS4A protein.

Production of an active NS31-181-NS4A peptide complex at present involves two steps. First, the NS3 catalytic domain (amino acid residues 1-181) is produced as a recombinant protein in E. coli. Next, a 13-19 residue NS4A peptide spanning the central hydrophobic domain of the full-length NS4A protein is added to form a non-covalent complex [Kim et al., 1996, Cell 87:343-355]. This complex, although more active than the protease alone, is approximately 8-10 fold less active than the full-length NS31-631-NS4A1-54 form of the protease as judged by its proteolytic activity toward a synthetic substrate based on the native NS5A-NS5B amino acid sequence. [Urbani et al., 1997, J. Biol. Chem., 272(14):9204-09; Steinkuhler et al., 1996, J. Virol. 70(10):6694-6700]. Moreover, NS4A peptide has been shown to have a very low affinity (10 uM) for NS3 in solution [Bianchi et al., 1997, Biochemistry 36: 7890-7897], requiring addition of NS4A peptide in the high micromolar range to insure a 1:1 stoichiometric complex with NS3 protease. The limited solubility of this peptide in aqueous buffer due to its hydrophobic nature makes working with this peptide at these concentrations difficult.

Because the HCV NS3 protease cleaves the non-structural HCV proteins necessary for HCV replication, the NS3 protease can be a target

15

35

for the development of therapeutic agents against the HCV virus. The gene encoding the HCV NS3 protein has been cloned as disclosed in U.S. Patent No. 5,371,017. To date, however, the protease has not been produced in a covalent complex with the NS4A cofactor in a soluble, active and stable form. Such a complex would be useful as a target in a high throughput screen to discover therapeutic agents. A stable, active HCV protease is also required for determination of modes of binding of inhibitors by NMR, for structural determination by NMR spectroscopy, for crystallography, and for virtually all biophysical and biochemical studies interested in the activated form of the enzyme.

SUMMARY OF THE INVENTION

The present invention provides NS4A tethered forms of the HCV NS3 protease comprising single-chain recombinant covalent complexes of Hepatitis C virus NS3 protease and an NS4A cofactor peptide which require no subsequent addition of NS4A peptide for activation and which are as active as the full-length NS31-631 NS4A1-54. The covalent 20 NS4A-NS3 complexes of the invention are more soluble, stable and active than the non-covalent protease-peptide complexes previously available.

The NS4A tethered forms of the HCV NS3 protease of the 25 invention consist of covalent NS4A-NS3 complexes comprising a central hydrophobic domain of the NS4A peptide tethered by linker of at least about 4 amino acid residues to the amino terminus of the serine protease domain of NS3. The amino acid sequences of 20 such embodiments are defined in the Sequence Listing by SEQ ID NOs: 1-20. 30 Corresponding nucleotide sequences are provided in SEQ ID NOs: 91-111.

Preferred embodiments of the invention also provide NS4A tethered forms of the full length NS3 protease. The amino acid sequences of 8 such embodiments are defined in SEQ ID NOs: 11-18.

Other preferred embodiments of the invention further provide mutant forms of the covalent NS4A-NS3 complexes in which point

10

15

20

mutations introduced at positions 17 and/or 18 of the NS3 domain change a hydrophobic amino acid residue to a hydrophilic residue. This further improves the solubility of the complexes and provides the protein in a monodispersed form. The amino acid sequences of 13 such embodiments are defined in the Sequence Listing by SEQ ID NOs: 2-4, 6-8, 10, 12-14, and 16-18.

The invention still further provides mutant forms of the covalent NS4A-NS3 complexes in which a mutation introduced at position 139 of the NS3 domain changes a serine residue to an alanine residue. The amino acid sequences of 9 such embodiments are defined in SEQ ID NOs: 5-8, 15-18 and 20.

The invention further provides covalent HCV NS4A-NS3 complexes having an easily removable histidine tag comprising three or more histidine residues fused to the complex. This enables rapid purification of the protease with easy removal of the tag following purification.

The present invention further provides for isolated nucleic acids and vectors which encode the covalent NS4A-NS3 complexes of the present invention, and host cells transformed or transfected by said nucleic acids or vectors.

25 The invention still further provides methods for making the covalent NS4A-NS3 complexes comprising culturing the transformed or transfected host cell under conditions in which the nucleic acid or vector is expressed.

30 The invention also provides methods for identifying inhibitors of HCV NS3. Methods are provided for detecting inhibitors of the protease activity, the helicase activity and the ATPase activity of NS3 using the disclosed covalent complexes.

35

BRIEF DESCRIPTION OF THE FIGURES

Figure 2 depicts the recombinant synthesis of plasmid pHIS-NS4A₂₁₋₃₂-GSGS-NS3₃₋₁₈₁

5 Figure 3 depicts the recombinant synthesis of plasmid pHIS-NS3₁₋₆₃₁.

Figure 4 depicts the recombinant synthesis of plasmid pHIS-NS4A₂₁₋₃₂-GSGS-NS3₃₋₆₃₁

10

15

Figures 5A and 5B schematically depict a high throughput assay for discovering HCV protease inhibitors using surface plasmon resonance technology. Figure 5A illustrates the outcome expected in the absence of an uninhibited HCV protease, while 5B illustrates the outcome expected in the presence of an active, uninhibited HCV protease.

Figure 6 shows the nucleic acid unwinding activity of the covalent His-NS4A₂₁₋₃₂-GSGS-NS₃₃₋₆₃₁ as compared to that of the His NS3₁₋₆₃₁/NS4A₁₋₅₄

20 Figure 7 shows the ATPase activity of the covalent His-NS4A₂₁₋₃₂-GSGS-NS₃₂₋₆₃₁ complex as monitored by thin layer chromatography.

DETAILED DESCRIPTION OF THE INVENTION

25

30

35

The teachings of all references cited are incorporated herein in their entirety by reference.

The covalent NS4A-NS3 complexes of the present invention are useful for structural determination and determination of mode of binding of HCV inhibitors by NMR spectroscopy. Moreover, they provide a more soluble and stable form of HCV NS3 protease than the presently available non-covalent NS3₁₋₁₈₁.NS4A peptide complexes for crystallography studies, high throughput screening assays and other conventional biophysical and biochemical investigations.

Several representative embodiments of the covalent NS4A-NS3 complexes of the invention are disclosed in the examples below. In one

15

20

25

30

such embodiment, NS4A residues 21-32 were tethered to the amino terminus of residues 3-181 of mature NS3 protease by a 4-residue linker, GSGS (SEQ ID NO: 21). The complex was overexpressed as a soluble protein in *E. coli* and purified to homogeneity by a combination of metal chelate and size-exclusion chromatography. The tethered complex, HIS-NS4A₂₁₋₃₂-GSGS-NS3₃₋₁₈₁ (SEQ ID NO: 1) cleaved a NS5A/5B synthetic substrate with a catalytic efficiency identical to that of the non-covalent full-length protease, NS3₁₋₆₃₁-NS4A₁₋₅₄.

In other embodiments of the invention, the NS4A hydrophobic domain and the NS3 serine protease domain are covalently tethered using different amino acid linkers. The preferred amino acid linkers of the invention comprise at least about four amino acid residues. More preferably, the linkers consist of from four to six amino acid residues. More preferably, four-residue linkers are used. Most preferably, amino acid linkers having the sequence defined by SEQ ID NO: 21 or 22 are used to tether the NS4A hydrophobic domain and the NS3 serine protease domain.

Routine procedures in the art would allow one to construct covalent NS4A-NS3 complexes of the invention having linkers of various sizes. It will be understood by one skilled in the art, for example, that if smaller or larger portions of the NS3 or NS4A domains are used to construct the covalent complexes of the invention, longer or shorter amino acid linkers can be used.

Other embodiments of the present invention contain smaller or larger portions of the NS4A cofactor peptide. In preferred embodiments, the complexes contain an NS4A hydrophobic domain comprising at least amino acid residues 22-30 of the full length NS4A cofactor peptide. More preferably, the complexes contain from 12-19 amino acid residues spanning the central hydrophobic domain of the full length NS4A peptide. Most preferably, the complexes contain amino acid residues 21-32 of full length NS4A peptide.

35

Still further embodiments of the present invention contain smaller or larger portions of the NS3 protease. In preferred embodiments, the complexes contain an NS3 serine protease domain

comprising at least amino acid residues 3-181 of the full length NS3 protease. More preferably, the complexes contain amino acid residues 1-181 of full length NS3 protease. Most preferably, the complexes contain amino acid residues 3-181 of full length NS3 protease.

5

10

The present invention thus also includes covalent NS4A-NS3 complexes comprising the central hydrophobic domain of the NS4A peptide tethered to the amino terminus of full-length mature NS3 protease (amino acids 1-631) by an amino acid linker. The amino acid sequences of preferred embodiments comprising NS4A tethered to full-length mature NS3 protease are set forth in SEQ ID NOs: 11-18.

15

20

25

Surprisingly, it has also been found that the introduction of point mutations at position 17 and/or 18 of the NS3 domain of the NS4A- NS3 constructs of the present invention which change a hydrophobic amino acid residue to a hydrophilic amino acid residue produces a more soluble and mono-dispersed form of the tethered complex. representative embodiments of such mutant NS4A-NS3 complexes are disclosed in the Examples below. In some embodiments, the isoleucine at position 17 is mutated to lysine. One such mutant form is referred to as His-NS4A₂₁₋₃₂-GSGS-NS3₃₋₁₈₁/I17K (SEQ ID NO: 2). embodiments, the same mutation is made at position 18. One such mutant form is referred to as His-NS4A21-32-GSGS-NS33-181/I18K (SEQ ID NO: 3). In yet other embodiments, the mutations are introduced at both positions. One such mutant is referred to as His-NS4A21-32-GSGS-NS33-181/I17K,I18K (SEQ ID NO: 4). Each of the purified mutants results in a monodispersed (as judged by size exclusion chromatography) and more soluble (as judged by achieving higher concentration of the complex 17-20 mg/ml) form of the complex, which remains monodispersed for a period of about one week at 4°C, while still exhibiting kinetic properties identical to those of the wild type.

30

35

It will be understood that although the foregoing embodiments are presently preferred, other modifications to the hydrophobic residues at positions 17 and 18 can be made to produce other soluble complexes. Preferably, neutral amino acid residues will be substituted for charged residues. These modifications can be used in a number of combinations to produce the final modified protein chain.

Also provided are NS4A-tethered forms of NS3 full-length domain. In contrast to the NS4A-tethered forms of the catalytic domain, a considerable amount of autocleavage in the helicase domain of the NS3 protein is detected during the purification of their native full-length counterpart, HIS-NS4A21-32-NS3₃₋₆₃₁. To prevent autocleavage of the full-length covalent complexes, the catalytic serine residue at position 139 is mutated to alanine. The amino acid sequence of one such embodiment is defined by SEQ ID NO: 15. The mutation of the full length constructs at position 139 can also be made in the NS4A-tethered forms of the NS3 catalytic domain, and can be made in combination with any of the aforementioned mutations to increase solubility and stability while preventing autocleavage. Representative embodiments are set forth in SEQ ID NOs: 5-8, 15-18 and 20.

15

20

10

As used herein, the terms "native NS3" and "full-length NS3" are used interchangeably and are defined as a protein which (a) has an amino acid sequence substantially identical to the sequence defined by SEQ ID NO: 23 and (b) has biological activity that is common to native NS3. This includes natural allelic variants and other variants having one or more conservative amino acid substitutions [Grantham, 1974, Science 185:862] that do not substantially impair biological activity. Such conservative substitutions involve groups of synonymous amino acids, e.g., as described in U.S. patent No. 5,017,691 to Lee et al.

25

The "serine protease domain" of NS3 or the "catalytic domain" of NS3 refers to amino acids 1-181 of mature NS3, which have been shown to contain the active catalytic triad His, Asp and Ser.

30

35

The term "native NS4A peptide" as used herein is defined as a peptide which (a) has an amino acid sequence substantially identical to the sequence defined by SEQ ID NO: 24; and (b) has biological activity that is common to native NS4A. This includes natural allelic variants and other variants having one or more conservative amino acid substitution [Grantham, 1974, Science 185:862] that do not substantially impair biological activity. Such conservative substitutions involve groups of synonymous amino acids, e.g., as described in U.S. patent No. 5,017,691 to Lee et al.

As used herein, the "central hydrophobic domain of NS4A peptide" refers to that portion of the native NS4A peptide (approximately amino acid residues 22 - 30) which is sufficient for activation of NS3 protease. Size and sequence variants of this domain which also activate the NS3 protease in the claimed complexes also fall within this term.

A "soluble" covalent complex as referred to herein is defined as a protein which will remain in solution after a high spin centrifugation step at $300,000 \times g$ in a standard ultracentrifuge in a buffer containing 25 mM HEPES, pH 7.6, 10% glycerol, 0.3 M NaCl, 10 mM β ME.

An "active" covalent complex as referred to herein is defined as a complex which will cleave synthetic substrates corresponding to NS5A-NS5B cleavage site (for example, DTEDVVCC SMYTWTGK) (SEQ ID NO: 25)) between P1 residue, cysteine and P1' residue, serine in a buffer containing 25 mM Tris, pH 7.5, 150 mM NaCl, 10 % glycerol, and 0.05 % lauryl maltoside.

20

25

30

35

10

15

Nucleic acids encoding the covalent NS4A-NS3 complexes are also a part of this invention. DNA encoding the covalent NS4A-NS3 complexes of this invention can be prepared by chemical synthesis using the known nucleic acid sequence [Ratner et al., 1985, Nucleic Acids Res. 13:5007] and standard methods such as phosphoramidite solid support method of Matteucci et al., 1981, J. Am. Chem. Soc. 103:3185 or the method of Yoo et al., 1989, J. Biol. Chem. 764:17078. See also Glick, Bernard R. and Pasternak, Molecular Biotechnology, pages 55 - 63, (ASM Press, Washington, D.C. 1994). The genes encoding the desired regions of the HCV protein can also be obtained using the plasmid disclosed in Grakoui, et al., 1993, J. Virol. 67:1385-1395 or that disclosed in Takamizawa et al., 1991, J. Virology 65(3):1105-1113. Also, the nucleic acid encoding HCV NS3 and NS4A can be isolated, amplified and cloned from patients infected with the HCV virus. Furthermore, the HCV genome has been disclosed in PCT WO 89/04669 and is available from the

American Type Culture Collection (ATCC), 12301 Parklawn Drive, Rockville, MD under ATCC accession no. 40394.

Of course, because of the degeneracy of the genetic code, there are many functionally equivalent nucleic acid sequences that can encode the NS3 and NS4A domains of the covalent NS4A-NS3 complexes as defined herein. Such functionally equivalent sequences, which can readily be prepared using known methods such as chemical synthesis, PCR employing modified primers and site-directed mutagenesis, are within the scope of this invention.

10

15

20

25

Various vectors can be used to express DNA encoding the covalent NS4A-NS3 complexes. Conventional vectors used for expression of recombinant proteins in prokaryotic or eukaryotic cells may be used. Preferred vectors include the pET vectors described by Studier et al., 1990, Methods of Enzymology 185: 60-89, and the pcD vectors described by Okayama et al., 1983, Mol. Cell. Bio. 3: 280-289; and Takebe et al., 1988, Mol. Cell. Biol. 8: 466-472. Other SV40-based mammalian expression vectors include those disclosed in Kaufman et al., 1982, Mol. Cell. Biol. 2: 1304-1319 and U.S. Patent No. 4,675,285. These SV40-based vectors are particularly useful in COS7 monkey cells (ATCC No. CRL 1651), as well as in other mammalian cells such as mouse L cells and CHO cells.

30

35

Standard transfection methods can be used to produce eukaryotic cell lines which express large quantities of polypeptides. Eukaryotic cell lines include mammalian, yeast and insect cell lines. Exemplary mammalian cell lines include COS-7 cells, mouse L cells and Chinese Hamster Ovary (CHO) cells. See Sambrook et al., supra and Ausubel et al., supra.

As used herein, the term "transformed bacteria" means bacteria that have been genetically engineered to produce a viral or mammalian protein. Such genetic engineering usually entails the introduction of an expression vector into a bacterium. The expression vector is capable of autonomous replication and protein expression relative to genes in the bacterial genome. Construction of bacterial expression vectors is well known in the art, provided the nucleotide sequence encoding a desired

15

20

25

30

35

protein is known or otherwise ascertainable. For example, DeBoer in U.S. Pat. No. 4,551,433 discloses promoters for use in bacterial expression vectors; Goeddel et al. in U.S. Pat. No. 4,601,980 and Riggs, in U.S. Pat. No. 4,431,739 disclose the production of mammalian proteins by E. coli expression systems; and Riggs supra, Ferretti et al., 1986, Proc. Natl. Acad. Sci. 83:599, Sproat et al., 1985, Nucleic Acid Research 13:2959 and Mullenbach et al., 1986, J. Biol. Chem 261:719 disclose how to construct synthetic genes for expression in bacteria. Many bacterial expression vectors are available commercially and through the American Type Culture Collection (ATCC), Rockville, Maryland.

Insertion of DNA encoding the covalent NS4A-NS3 complexes into a vector is easily accomplished when the termini of both the DNA and the vector comprise the same restriction site. If this is not the case, it may be necessary to modify the termini of the DNA and/or vector by digesting back single-stranded DNA overhangs generated by restriction endonuclease cleavage to produce blunt ends, or to achieve the same result by filling in the single-stranded termini with an appropriate DNA polymerase.

Alternatively, any site desired may be produced by ligating nucleotide sequences (linkers) onto the termini. Such linkers may comprise specific oligonucleotide sequences that define desired restriction sites. The cleaved vector and the DNA fragments may also be modified if required by homopolymeric tailing.

Many *E. coli*-compatible expression vectors can be used to produce soluble covalent NS4A-NS3 complexes of the present invention, including but not limited to vectors containing bacterial or bacteriophage promoters such as the Tac, Lac, Trp, LacUV5, λ P_r and λ P_L promoters. Preferably, a vector selected will have expression control sequences that permit regulation of the rate of expression. Then, production of covalent NS4A-NS3 complexes can be regulated to avoid overproduction that could prove toxic to the host cells. Most preferred is a vector comprising, from 5' to 3' (upstream to downstream), a Tac promoter, a Iac II repressor gene and DNA encoding mature human HCV protease. The vectors chosen for use in this invention may also encode secretory leaders such as the ompA or protein A leader, as long as such leaders are cleaved during

post-translational processing to produce covalent NS4A-NS3 complexes or if the leaders are not cleaved, the leaders do not interfere with the enzymatic activity of the protease.

The covalent complexes of the invention, or portions thereof, can also be synthesized by a suitable method such as by exclusive solid phase synthesis, partial solid phase methods, fragment condensation or classical solution synthesis. The polypeptides are preferably prepared by solid phase peptide synthesis as described by Merrifield, 1963, J. Am. Chem. Soc. 85:2149. The synthesis is carried out with amino acids that are protected at the alpha-amino terminus. Trifunctional amino acids with labile side-chains are also protected with suitable groups to prevent undesired chemical reactions from occurring during the assembly of the polypeptides. The alpha-amino protecting group is selectively removed to allow subsequent reaction to take place at the amino-terminus. The conditions for the removal of the alpha-amino protecting group do not remove the side-chain protecting groups.

The alpha-amino protecting groups are those known to be useful in the art of stepwise polypeptide synthesis. Included are acyl type protecting groups (e.g., formyl, trifluoroacetyl, acetyl), aryl type protecting groups (e.g., biotinyl), aromatic urethane type protecting groups [e.g., benzyloxycarbonyl (Cbz), substituted benzyloxycarbonyl and 9-fluorenylmethyloxy-carbonyl (Fmoc)], aliphatic urethane protecting groups ſe.g., t-butyloxycarbonyl (tBoc). isopropyloxycarbonyl, cyclohexyloxycarbonyl] alkvl protecting groups (e.g., benzyl, triphenylmethyl). The preferred protecting groups are tBoc and Fmoc, thus the peptides are said to be synthesized by tBoc and Fmoc chemistry, respectively.

30

35

25

5

10

15

20

The side-chain protecting groups selected must remain intact during coupling and not be removed during the deprotection of the amino-terminus protecting group or during coupling conditions. The side-chain protecting groups must also be removable upon the completion of synthesis, using reaction conditions that will not alter the finished polypeptide. In tBoc chemistry, the side-chain protecting groups for trifunctional amino acids are mostly benzyl based. In Fmoc chemistry, they are mostly tert.-butyl or trityl based.

15

20

25

30

35

In tBoc chemistry, the preferred side-chain protecting groups are tosyl for Arg, cyclohexyl for Asp, 4-methylbenzyl (and acetamidomethyl) for Cys, benzyl for Glu, Ser and Thr, benzyloxymethyl (and dinitrophenyl) for His, 2-Cl-benzyloxycarbonyl for Lys, formyl for Trp and 2-bromobenzyl for Tyr. In Fmoc chemistry, the preferred side-chain protecting groups are 2,2,5,7,8-pentamethylchroman-6-sulfonyl (Pmc) or 2,2,4,6,7-pentamethyldihydrobenzofuran-5-sulfonyl (Pbf) for Arg, trityl for Asn, Cys, Gln and His, tert butyl for Asp, Glu, Ser, Thr and Tyr, tBoc for Lys and Trp.

For the synthesis of phosphopeptides, either direct or postassembly incorporation of the phosphate group is used. In the direct incorporation strategy, the phosphate group on Ser, Thr or Tyr may be protected by methyl, benzyl or tert.butyl in Fmoc chemistry or by methyl, benzyl or phenyl in tBoc chemistry. Direct incorporation of phosphotyrosine without phosphate protection can also be used in Fmoc chemistry. In the post-assembly incorporation strategy, the unprotected hydroxyl group of Ser, Thr or Tyr is derivatized on solid with di-tert.butvl-. dibenzylordimethyl-N,N'diisopropylphosphoramidite and then oxidized by tert.butylhydroperoxide.

Solid phase synthesis is usually carried out from the carboxylterminus by coupling the alpha-amino protected (side-chain protected) amino acid to a suitable solid support. An ester linkage is formed when the attachment is made to a chloromethyl, chlortrityl or hydroxymethyl resin, and the resulting polypeptide will have a free carboxyl group at the C-terminus. Alternatively, when an amide resin such as benzhydrylamine or p-methylbenzhydrylamine resin (for tBoc chemistry) and Rink amide or PAL resin (for Fmoc chemistry) is used, an amide bond is formed and the resulting polypeptide will have a carboxamide group at the C-terminus. These resins. whether polystyreneor polvamide-based polyethyleneglycol-grafted, with or without a handle or linker, with or without the first amino acid attached, are commercially available, and their preparations have been described by Stewart et al (1984)., "Solid Phase Peptide Synthesis" (2nd Edition), Pierce Chemical Co., Rockford, IL.; and Bayer & Rapp (1986) Chem. Pept. Prot. 3, 3; and Atherton, et al. (1989) Solid Phase Peptide Synthesis: A Practical Approach, IRL Press, Oxford.

5

10

15

The C-terminal amino acid, protected at the side-chain if necessary and at the alpha-amino group, is attached to a hydroxylmethyl resin using various activating agents including dicyclohexylcarbodiimide (DCC), N,N'-diisopropylcarbodiimide DIPCDI) and carbonyldiimidazole (CDI). It can be attached to chloromethyl or chlorotrityl resin directly in its cesium tetramethylammonium salt form or in the presence of triethylamine (TEA) or diisopropylethylamine (DIEA). First amino acid attachment to an amide resin is the same as amide bond formation during coupling reactions.

Following the attachment to the resin support, the alphaamino protecting group is removed using various reagents depending on the protecting chemistry (e.g., 180c, Fmoc). The extent of Fmoc removal can be monitored at 300-320 nm or by a conductivity cell. After removal of the alpha-amino protecting group, the remaining protected amino acids are coupled stepwise in the required order to obtain the desired sequence.

25

30

35

20

Various activating agents can be used for the coupling reactions including DCC, DIPCDI, 2-chloro-1,3-dimethylimidium hexafluorophosphate (CIP), benzotriazol-1-yl-oxy-tris-(dimethylamino)-phosphonium hexafluorophosphate (BOP) and its pyrrolidine analog (PyBOP), bromo-tris-pyrrolidino-phosphonium hexafluorophosphate (PvBroP). O -(benzotriazol-1-vl)-1.1.3.3tetramethyluronium hexafluorophosphate (HBTU) tetrafluoroborate analog (TBTU) or its pyrrolidine analog (HBPyU), -(7-azabenzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HATU) and its tetrafluoroborate analog (TATU) or pyrrolidine analog (HAPyU). The most common catalytic additives coupling reactions include dimethylaminopyridine (DMAP), 3-hydroxy-3,4-dihydro-4-oxo-1,2,3benzotriazine (HODhbt), N-hydroxybenzotriazole (HOBt) and 1-

15

20

25

30

35

hydroxy-7-azabenzotriazole (HOAt). Each protected amino acid is used in excess (>2.0 equivalents), and the couplings are usually carried out in N-methylpyrrolidone (NMP) or in DMF, CH2Cl2 or mixtures thereof. The extent of completion of the coupling reaction can be monitored at each stage, e.g., by the ninhydrin reaction as described by Kaiser et al., Anal. Biochem. 34:595 (1970). In cases where incomplete coupling is found, the coupling reaction is extended and repeated and may have chaotropic salts added. The coupling reactions can be performed automatically with commercially available instruments such as ABI model 430A, 431A and 433A peptide synthesizers.

After the entire assembly of the desired polypeptide, the polypeptide-resin is cleaved with a reagent with proper scavengers. The Fmoc peptides are usually cleaved and deprotected by TFA with scavengers (e.g., H2O, ethanedithiol, phenol and thioanisole). The tBoc peptides are usually cleaved and deprotected with liquid HF for 1-2 hours at -5 to 0°C, which cleaves the polypeptide from the resin and removes most of the side-chain protecting groups. Scavengers such as anisole, dimethylsulfide and p-thiocresol are usually used with the liquid HF to prevent cations formed during the cleavage from alkylating and acylating the amino acid residues present in the polypeptide. The formyl group of Trp and dinitrophenyl group of His need to be removed, respectively, by piperidine and thiophenol in DMF prior to the HF cleavage. The acetamidomethyl group of Cys can be removed by mercury(II) acetate and alternatively by iodine, thallium (III) trifluoroacetate or silver tetrafluoroborate which simultaneously oxidize cysteine to cystine. Other strong acids used for tBoc peptide cleavage and deprotection include trifluoromethanesulfonic acid (TFMSA) and trimethylsilyltrifluoroacetate (TMSOTf).

Recombinant DNA methodology can also be used to prepare the polypeptides. The known genetic code, tailored if desired with known preferred codons for more efficient expression in a given host organism, can be used to synthesize oligonucleotides encoding the desired amino acid sequences. The phosphoramidite solid support method of Matteucci et al., J. Am. Chem. Soc. 103:3185 (1981) or other

known methods can be used for such syntheses. The resulting oligonucleotides can be inserted into an appropriate vector and expressed in a compatible host organism.

The polypeptides of the invention can be purified using HPLC, gel filtration, ion exchange and partition chromatography, countercurrent distribution or other well known methods. In a preferred embodiment of the present invention the covalent NS4A-NS3 complexes also contain a histidine tag which facilitates purification using a Ni⁺ column as is illustrated below.

One can use the covalent NS4A-NS3 complexes of the invention, along with known synthetic substrates, to develop high throughput assays. These can be used to screen for compounds which inhibit proteolytic activity of the protease. This is carried out by developing techniques for determining whether or not a compound will inhibit the covalent NS4A-NS3 complexes of the invention from cleaving the viral substrates. Examples of such synthetic substrates are set forth in SEQ ID NOs 25 and 93. If the substrates are not cleaved, the virus cannot replicate. One example of such a high throughput assay is the scintillation proximity assay (SPA). SPA technology involves the use of beads coated with scintillant. Bound to the beads are acceptor molecules such as antibodies, receptors or enzyme substrates which interact with ligands or enzymes in a reversible manner.

25

30

20

5

10

15

For a typical protease assay the substrate peptide is biotinylated at one end and the other end is radiolabelled with low energy emitters such as ¹²⁵I or ³H. The labeled substrate is then incubated with the enzyme. Avidin coated SPA beads are then added which bind to the biotin. When the substrate peptide is cleaved by the protease, the radioactive emitter is no longer in proximity to the scintillant bead and no light emission takes place. Inhibitors of the protease will leave the substrate intact and can be identified by the resulting light emission which takes place in their presence.

35

Another type of protease assay, utilizes the phenomenon of surface plasmon resonance (SPR). A novel, high throughput enzymatic assay utilizing surface plasmon resonance technology has been

successfully developed. Using this assay, and a dedicated BIAcoreTM instrument, at least 1000 samples per week can be screened for either their enzymatic activity or their inhibitory effects toward the enzymatic activity, in a 96 well plate format. This methodology is readily adaptable to any enzyme-substrate reaction. The advantage of this assay over the SPA assay is that it does not require a radiolabeled peptide substrate.

15

20

25

30

EXAMPLES

Several covalent NS4A-NS3 complexes have been constructed, purified, characterized and assayed for activity based on a cDNA clone containing an HCV Japanese (1b/BK) strain whose sequence is published in Takamizawa *et al.*, 1991, *J. Virology* 65:1105-1113. DNA sequencing of the clone (BK 138-1) revealed four amino acid differences with the published sequence, at positions 66 (A->G), 86 (P->Q), 87 (K->A) and 147 (F->S) of the NS3 protein.

The present invention can be illustrated by the following nonlimiting examples.

Reagents and General Methods

Plasmid pHCV-1b/BK can be derived from DNA fragments containing the entire DNA sequence of HCV BK cDNA as reported by Takamizawa *et al.*, 1991, *J. Virology* 65:1105-1113, with the abovementioned changes. Plasmid pMD-34-2 is derived from that portion of the disclosed DNA sequence which encodes NS3 residues 1-631 from HCV BK cDNA.

Restriction Enzymes, Vent Polymerase and ThermoPol buffer were obtained from New England Biolabs (Beverly, MA). The QuickChange mutagenesis kit and dNTP's were obtained from Stratagene (LaJolla, CA). Ready-to-Go T4 DNA Ligase was obtained from Pharmacia Biotech (Piscataway, NJ). Oligonucleotide primers were synthesized by Genosys Biotechnologies (Woodland, Texas). DNA sequencing was performed according to the Sanger-Dideoxy method by Bioserve Biotechnologies (Laurel, MD). pET vectors and BL21(DE3) cells were obtained from Novagen (Madison, WI). PCR reactions were carried out in a Perkin Elmer Cetus, model 480 DNA thermocycler. DH5 α cells and TAE buffer were purchased from Gibco, BRL. GTG agarose was purchased from FMC corporation. The Qiaquick gel extraction kit and Qiaquick PCR purification kit were purchased from Qiagen Inc. (Chatsworth, CA).

15

20

25

30

Standard DNA recombinant DNA methods were carried out essentially as described by Sambrook et. al. in "Molecular Cloning: A Laboratory Manual," 2nd edition, 1989, Cold Springs Harbor Press, Plainview, New York.

5 Preparation of NS4A-Tethered Forms of HCV NS3 Protease

Native, NS4A-tethered forms of NS3 catalytic domain

Various NS4A-tethered forms of the NS3 catalytic domain were constructed by joining the NS4A peptide GSVVIVGRIILS (NS4A amino acids 21-32) to the amino terminus of NS3 amino acids 3-181 via various three or four residue linkers, and were cloned into the pET-28b+vector.

Single stranded oligonucleotide primers were designed to generate a 616 base pair PCR fragment containing an NdeI site followed by the NS4A peptide, a linker, and amino acids 3-181 of the NS3 catalytic domain at the 5' terminus and a stop codon flanked by an EcoRI site at the 3' terminus. The template used was the sequence disclosed in Takamizawa, et al, 1991, J. Virology 65(3):1105-1113, which contains the entire HCV genome from the 1b/BK strain, except for the four differences described above. Other sources for HCV DNA can be used in the disclosed methods, including plasmid pBRTM/HCV 1-3011 (Grakoui et al., 1993), which contains the entire genome from the 1a strain.

Vent DNA polymerase was utilized to amplify the DNA by PCR. Primers were diluted in dH_20 to give a final concentration of 50 μ g/ml. The template was diluted in dH_20 to give a final concentration of 10 ng/ μ l; The dNTP's (GTP, ATP, CTP, GGT) were diluted to a concentration of 10 mM (2.5 mM each) in dH_20 .

100 μ l reactions were prepared for PCR in a 500 ul Eppendorf tube by addition of the following reagents: 74 μ l of dH20, 10 ul of the 10x Thermopol buffer (final 1x buffer: 10 mM KCL, 20 mMTris-HCL (pH 8.8), 2mM MgSO₄ and 0.1% Triton X), 10 μ l of template (100 ng), 2 μ l of the 5′ primer (100 ng); 1 μ l of the 3′ primer (50 ng), 2 μ l of the dNTP mixture (200 μ M) and 1 μ l of Vent polymerase enzyme (1 unit). The mixture was

10

15

20

25

30

35

then overlayed with 20 ul of immersion oil and placed in the thermocycler for amplification. The PCR conditions were as follows: 95 $^{\circ}$ C for 45 seconds (1 cycle); 95 $^{\circ}$ C for 30 seconds, 55 $^{\circ}$ C for 1 minute, 72 $^{\circ}$ C for 2 minutes (25 cycles).

The amplified 616 base pair fragment was purified in preparation for restriction digestion using a Qiaquick PCR purification kit according to the manufacturer's protocol without modification. aqueous layer was removed and placed in a 1.5 ml Eppendorf tube with a regent that aids the DNA to bind to a column matrix. The DNA was washed while bound to the column and then eluted with 43 µl of H20. The DNA was then double digested with EcoRI and NdeI in a 50 ul volume for 1 hour at 37 °C. The reaction took place in a 1.5 ml polypropylene Eppendorf tube with 5 µl of 10x EcoRI buffer (final concentration of 50mM NaCl, 100 mM Tris-HCL, 10mM MgCl₂, 0.25% Triton X-100, pH 7.5) and $\mu 1$ 1 of EcoRI and NdeI (20 units). The pET-28b+ vector (3 µg) was also digested using the same conditions. The digests were further purified by resolving them on a 1.0 % agarose electrophoresis gel for 45 minutes under 100 volts. They were rendered visible with 0.5 $\mu g/ml$ of ethidium bromide, excised with a scalpel under short-wave UV, solubilized and purified using the QIAquick gel extraction kit according to manufacturer's protocol without modifications. The fragments were quantitated by visually comparing a 5 ul aliquot of the purified fragment versus Lambda Hind/III DNA standards on a 1% agarose gel. Approximately 200 ng of vector and 50 ng of PCR fragment were ligated together in a 20 ul volume for 18 hours at 16 degrees. They were combined together in a T4 ligase (Ready-to-Go) reaction tube according to standard protocol without modifications.

 $2~\mu l$ of this mixture was then used to transform $50~\mu l$ of DH5 α cells for plasmid propagation according to manufacturer's protocol. Briefly, a 1.5 ml Eppendorf tube was placed on ice and 50 ul of DH5 α cells (previously stored at -80°C and then thawed on ice immediately prior to use) were added to the tube along with the 2 ul of ligation mixture and allowed to incubate for 30 minutes. They were then heat shocked for 1 minute at 42°C, returned to the ice for 2 minutes and then regenerated with 500 μl of SOC medium and incubated at 37°C for 1 hour at 300 rpm.

15

20

25

30

35

 $200 \ \mu l$ of these cells were then plated out on LB/20-10-5 agar (per liter: tryptone 50 grams, yeast extract 25 grams, NaCl 12.5 gram) with kanamycin (25 μ g/ml), spread for single colony isolation and incubated at 37 °C overnight. Three single colonies were selected for plasmid preparations. They were inoculated into 100 mls of LB/20-10-5 broth with kanamycin (25 $\mu g/ml$) in a 250 ml baffled flask and grown overnight for 18 hours at 37 degrees at 300 RPM in a shaker. The next day, the cultures were spun down in 500 ml Nalgene centrifuge bottles (8000 RPM, 10 minutes, 4 °C) and the pellet was harvested for plasmid The Qiagen midi-prep kit was used according to isolation. manufacturer's protocol. The DNA was quantitated using a UV/VIS spectrophotometer (Perkin-Elmers) at 260 nm. The purified, plasmid-DNA isolates were sequenced on an Applied Biosystems 373A DNA sequencer at Bioserve Biotechnologies, Inc. To confirm the sequence, both top and bottom strands were sequenced via primers that were synthesized by Bioserve Biotechnologies.

Native, NS4A-tethered forms of NS3 full-length domain

Both parental plasmids, HIS-NS4A $_{21-32}$ -GSGS-NS3 $_{3-631}$ and HIS-NS4A $_{21-32}$ -GSGS-NS3 $_{3-631}$ /S139A parental plasmids were created via a cut and paste method. Briefly, 5 μ l of plasmid PMD34-2 (1 μ g), plasmid HIS-NS4A $_{21-32}$ -GSGS-NS3 $_{3-181}$ (5 μ g) and plasmid HIS-NS31 $_{1-631}$ /S139A (1 μ g) were each digested separately in a 1.5 ml Eppendorf tube with 5 μ l of NEB buffer #2 (at final concentration of 10mM Tris-HCL, 10mM MgCl $_2$, 50mM NaCl, 1mM DTT, pH 7-9), 0.5 μ l of acetylated BSA (final concentration 100 μ g/ml), 1 μ l of XbaI (2 Units) and 38.5 μ l of ddH $_2$ 0.

These digests were incubated at 37 °C for one hour at which time 2.5 μ l of 2M NaCl (final concentration of 150mM) 45 μ l of ddH₂0 and 2.5 μ l of BspMI (2 Units) were added to the digests and incubated for 2 more hours at 37 °C. The double digests were then resolved on 0.8 % agarose gels and the size and quantity of the fragments were determined. The agarose gels were electorphoresed in BioRad apparatus and the fragments were excised using a scalpel. The excised backbone fragments which were derived from PMD34-2 and HIS-NS3₁₋₆₃₁/S139A were each 7.1 KB and the insert from HIS-NS4A₂₁₋₃₂-CSGS-NS3₃₋₁₈₁ was 275 base pairs. Approximately 2 μ l of 7.1 KB backbone (200 ng) and 1 μ l of 225 bp

15

20

25

30

35

insert (50 ng) were ligated together in a 20 μ l volume for 18 hours at 16 °C. They were combined together in a T4 ligase (Ready-to-Go) reaction tube according to standard protocol without modifications. 2 μ l of this mixture was then used to transform 50 μ l of DH5 α cells for plasmid propagation according to manufacturer's protocol.

Three single colonies of each construct were selected for miniprep plasmid isolations using a Qiagen miniprep kit. They were inoculated into 5 mls of LB/20-10-5 broth with ampicillin (100 $\mu g/ml$) in a 15 ml tubes and grown overnight for 18 hours at 37°C at 300 RPM in a shaker. The next day, the cultures were spun down 3000 RPM, 10 minutes, 4°C and the pellet was harvested for plasmid isolation. The clones were then assessed for recombination by digesting with BspMI and Xba1 according to the conditions described above. The digests were resolved on a 1% agarose gel and only those constructs yielding a 225 bp and 7.1 KB bp fragment were chosen as positives. Cultures from the positive clones were inoculated into 100 mls of LB/20-10-5 broth with ampicillin (100 ug/ml) in a 250 ml baffled flask and grown overnight for 18 hours at 37°C at 300 RPM in a shaker. The next day, the cultures were spun down in 500 ml Nalgene centrifuge bottles (8000 RPM, 10 minutes, 4°C) and the pellet was harvested for plasmid isolation. The Qiagen midiprep kit was used according to manufacturer's protocol. The DNA was quantitated using a UV/VIS spectrophotometer (Perkin-Elmers) at 260 The purified plasmid-DNA isolates were sequenced at the restriction site junctions on an Applied Biosystems 373A DNA sequencer at Bioserve Biotechnologies, Inc.

Site-directed Mutants

All site-directed mutations created in either NS4A-tethered forms of catalytic or full-length domain of NS3 protease were carried out using the quikchange site-directed mutagenesis kit (Stratagene) according to the manufacturer's protocol. For each mutation, two oligonucleotide primers (10 picomoles each) containing the desired mutation were used to amplify the entire plasmid encompassing the NS4A-tethered NS3 protease gene (50 or 100 ng/reaction) using pfu DNA polymerase (2.5 units/reaction) in a final reaction volume of 50 µl. The PCR conditions were as follows: 95 °C for 45 seconds (1 cycle); 95 °C for 30 seconds, 55 °C

15

25

for 1 minute, 68 °C for 15 minutes (16 cycles). After amplification, the reaction mixture was treated with 1 ul of DpnI (1 Unit) for 1 hour at 37 °C in order to digest the parental DNA.

One microliter of this digest was used to transform 50 μ l of XLI Blue cells to repair nicks and propagate the mutated plasmid. Plasmid-DNA were purified and transformed into BL21 (DE3) cells for expression studies.

EXAMPLE 1

NS3 Catalytic Domain Constructs

i. HIS-NS4A₂₁₋₃₂-GSGS-NS3₃₋₁₈₁ (SEQ ID NO: 1)

HIS-NS4A $_{21-32}$ -GSGS-NS3 $_{3-181}$ was constructed by joining amino acids 21-32 of the NS4A peptide to the N-terminal domain of NS3 protease (NS3 amino acids 3-181) via the linker GSGS (SEQ ID NO: 21), and was cloned into the pET-28b+ vector as described above. The 5′ primer reads as follows:

5'GATATACATATGGGTTCTGTTGTTATTGTTGGTAGAATTATTTTATCT GGTAGTGGTAGTATCACGGCCTACTCCCAA 3' (SEQ ID NO:26).

The 3' primer reads as follows:

20 5' CTCAGCGAATTCTCAAGACCGCATAGTAGTTTCCAT 3' (SEQ ID NO:27).

ii. HIS-NS4A₂₁₋₃₂-GSGS-NS3₃₋₁₈₁/I17K (SEQ ID NO: 2)

A single amino acid mutant of HIS-NS4A₂₁₋₃₂-GSGS-NS3₃₋₁₈₁ was constructed by creating a point mutation at position 17 of the NS3 domain of HIS-NS4A₂₁₋₃₂-GSGS-NS3₃₋₁₈₁ construct as described above. Two oligonucleotide primers, each complementary to opposite strands of the template weregenerated which contain the point mutation which

20

alters amino acid number 17 (isoleucine) to a lysine. The top strand primer was as follows:

5'CGGGGCCTACTTGGTTGCAAGATCACTAGCCTTACAGGC3'

5 (SEQ ID NO:28).

The bottom strand read as follows:

5' GCCTGTAAGGCTAGTGATCTTGCAACCAAGTAGGCCCCG 3'

(SEQ ID NO: 29).

10 The template, HIS-NS4A₂₁₋₃₂-GSGS-NS3₃₋₁₈₁, along with these two primers, were utilized in a PCR reaction to generate the point mutation.

(iii) HIS-NS4A₂₁₋₃₂-GSGS-NS3₃₋₁₈₁/I18K (SEQ ID NO: 3)

A single amino acid mutant of HIS-NS4A₂₁₋₃₂-GSGS-NS3₃₋₁₈₁ was constructed by creating a point mutation at position 18 of the NS3 domain of HIS-NS4A₂₁₋₃₂-GSGS-NS3₃₋₁₈₁ construct as described above. Two oligonucleotide primers, each complementary to opposite strands of the template, were generated which contain the point mutation which alters amino acid number 18 (isoleucine) to a lysine. The top strand primer was as follows:

5' CGGGGCCTACTTGGTTGCATCAAGACTAGCCTTACAGGC 3'

(SEQ ID NO: 30).

The bottom strand reads as follows:

5' GCCTGTAAGGCTAGTCTTGATGCAACCAAGTAGGCCCCG 3'

25 (SEQ ID NO: 31).

The template, HIS-NS4A $_{21-32}$ -GSGS-NS3 $_{3-181}$, along with these two primers was utilized in a PCR reaction to generate the point mutation.

(iv) HIS-NS4A₂₁₋₃₂-GSGS-NS3₃₋₁₈₁/I17K, I18K (SEQ ID NO: 4)

20

A single amino acid mutant of HIS-NS4A₂₁₋₃₂-GSGS-NS3₃₋₁₈₁/118K was constructed by creating a point mutation at position 17 of the NS3 domain of HIS-NS4A₂₁₋₃₂-GSGS-NS3₃₋₁₈₁/I18K construct as described above. Two oligonucleotide primers, each complementary to opposite strands of the template, were generated which contain the point mutation which alters amino acid number 18 (isoleucine) to a lysine. The top strand primer was as follows:

5' CGGGGCCTACTTGGTTGCAAGAAGACTAGCCTTACAGGC 3' (SEQ ID NO:32).

10 The bottom strand read as follows:

 $5'\ GCCTGTAAGGCTAGTCTTCTTGCAACCAAGTAGGCCCCG\ 3'.$

(SEQ ID NO:33)

The template HIS-NS4A $_{21:32}$ -GSGS-NS3 $_{3-181}$ /I18K, along with these two primers, was utilized in a PCR reaction to generate the point mutation.

v. HIS-NS4A₂₁₋₃₂-GSGS-NS3₃₋₁₈₁/S139A (SEQ ID NO: 5)

A single amino acid mutant of HIS-NS4A₂₁₋₃₂-GSGS-NS3₃₋₁₈₁ was constructed by creating a point mutation at position 139 of the NS3 domain of HIS-NS4A₂₁₋₃₂-GSGS-NS3₃₋₁₈₁ construct as described above. Two oligonucleotide primers, each complementary to opposite strands of the template, were generated which contain the point mutation which alters amino acid number 139 (catalytic serine) to an alanine. The top strand primer was as follows:

5' CTCCTACTTGAAGGGCTCTGCTGGTGGTCCACTGCTCTGC 3'

25 (SEO ID NO:34).

The bottom strand reads as follows:

5' GCAGAGCAGTGGACCACCAGCAGAGCCCTTCAAGTAGGAG 3' (SEQ ID NO:35).

The template HIS-NS4A_{21.32}-GSGS-NS3₃₋₁₈₁, along with these two primers, was utilized in a PCR reaction to generate the point mutation.

vi. HIS-NS4A₂₁₋₃₂-GSGS-NS3₃₋₁₈₁/S139A, I17K (SEQ ID NO: 6)

A single amino acid mutant of HIS-NS4A $_{21.32}$ -CSGS-NS3 $_{3.1}$ /S139A was constructed by creating a point mutation at position 17 of the NS3 domain of HIS-NS4A $_{21.32}$ -CSGS-NS3 $_{3.181}$ / S139A construct as described above. Two oligonucleotide primers, each complementary to opposite strands of the template, were generated which contain the point mutation which alters amino acid number 17 (isoleucine) to a lysine. The top strand primer was as follows:

5' CGGGGCCTACTTGGTTGCAAGATCACTAGCCTTACAGGC 3'

(SEQ ID NO:36).

The bottom strand reads as follows:

5' GCCTGTAAGGCTAGTGATCTTGCAACCAAGTAGGCCCCG 3'

15 (SEQ ID NO:37).

The template, HIS-NS4A $_{21-32}$ -GSGS-NS3 $_{3-181}$ /S139A, along with these two primers, was utilized in a PCR reaction to generate this point mutation.

vii. HIS-NS4A₂₁₋₃₂-GSGS-NS3₃₋₁₈₁/S139A, I18K (SEQ ID NO: 7)

A single amino acid mutant of HIS-NS4A₂₁₋₃₂-GSGS-NS3₃20 ₁₈₁/S139A was constructed by creating a point mutation at position 18 of
the NS3 domain of HIS-NS4A₂₁₋₃₂-GSGS-NS3₃₋₁₈₁/S139A construct as
described above. Two oligonucleotide primers, each complementary to
opposite strands of the template, were generated which contain the point
mutation which alters amino acid number 18 (isoleucine) to a lysine.

25 The top strand primer was as follows:

5' CGGGGCCTACTTGGTTGCATCAAGACTAGCCTTACAGGC 3'

(SEQ ID NO:38).

The bottom strand reads as follows:

5' GCCTGTAAGGCTAGTCTTGATGCAACCAAGTAGGCCCCG 3'

(SEQ ID NO:39).

The template, HIS-NS4A₂₁₋₃₂-GSGS-NS3₃₋₁₈₁/S139A along with these two primers was utilized in a PCR reaction to generate this point mutation.

5 viii. HIS-NS4A₂₁₋₃₂-GSGS-NS3₃₋₁₈₁/S139A, I17K, I18K (SEQ ID NO. 8)

A single amino acid mutant of HIS-NS4A $_{21.32}$ -GSGS-NS3 $_{3.181}$ /S139A, I17K was constructed by creating a point mutation at position 18 of the NS3 domain of HIS-NS4A $_{21.32}$ -GSGS-NS3 $_{3.181}$ /S139A,I17K construct as described above. Two oligonucleotide primers, each complementary to opposite strands of the template were generated which contain the point mutation which alters amino acid number 18 (isoleucine) to a lysine. The top strand primer was as follows:

5' CGGGGCCTACTTGGTTGCAAGAAGACTAGCCTTACAGGC 3'

(SEQ ID NO: 40).

15 The bottom strand reads as follows:

5' GCCTGTAAGGCTAGTCTTGATGCAACCAAGTAGGCCCCG 3'

(SEQ ID NO: 41).

The template, HIS-NS4A₂₁₋₃₂-GSGS-NS3₃₋₁₈₁/S139A,I17K, along with these two primers, was utilized in a PCR reaction to generate this point mutation.

ix. HIS-NS4A₂₁₋₃₂-PAGG-NS3₃₋₁₈₁ (SEQ ID NO: 9)

An NS4A-tethered form of the NS3 catalytic domain, HIS-NS4A₂₁.

32-PAGG-NS3₃₋₁₈₁, was constructed by joining the NS4A peptide GSVVIVGRIILS (NS4A amino acids 21-32) to the N-terminal domain of NS3 protease (NS3 amino acids 3-181) via the linker PAGG (SEQ ID NO: 22), and was cloned into the pET-28b+ vector as described above. Primers were designed to generate a 616 base pair PCR fragment containing an Ndel site followed by the NS4A peptide, the PAGG linker, and amino acids 3-181 of the NS3 catalytic domain at the 5' terminus and a stop

15

codon flanked by an EcoRI site at the 3' terminus. The 5' primer reads as follows:

5' GATATACATATGGGTTCTGTTGTTATTGTTGGTAGAATTATTTT

ATCTCCTGCTGGTGGTATCACGGCCTACTCCCAA 3' (SEQ ID NO: 42).

5 The 3' primer reads as follows:

 5^{\prime} CTCAGCGAATTCTCAAGACCGCATAGTAGTTTCCAT 3^{\prime} (SEQ ID NO: 43).

Plasmid-DNA (PMD-34-2), comprised of pET-22b+ vector encompassing the gene insert encoding HIS-NS3 (1-631) from 1b/BK strain was used as the template for PCR.

x. HIS-NS4A₂₁₋₃₂-PAGG-NS3₃₋₁₈₁/I17K (SEQ ID NO: 10)

A single amino acid mutant of HIS-NS4A $_{21\cdot32}$ -PAGG-NS3 $_{3\cdot181}$ was constructed by creating a point mutation at position 17 of the NS3 domain of the HIS-NS4A $_{21\cdot32}$ -PAGG-NS3 $_{3\cdot181}$ construct as described above. Two oligonucleotide primers, each complementary to opposite strands of the template, were generated which contain the point mutation which alters amino acid number 17 (isoleucine) to a lysine. The top strand primer was as follows:

5' CGGGGCCTACTTGGTTGCAAGATCACTAGCCTTACAGGC 3'

20 (SEQ ID NO: 44).

The bottom strand reads as follows:

5' GCCTGTAAGGCTAGTGATCTTGCAACCAAGTAGGCCCCG 3' (SEQ ID NO: 45).

The template, HIS-NS4A $_{21-32}$ -PAGG-NS3 $_{3-181}$, along with these two primers was utilized in a PCR reaction to generate this point mutation.

xi. HIS-NS4A₂₁₋₃₂-PAG-NS3₃₋₁₈₁ (SEQ ID NO: 46)

A NS4A-tethered form of the NS3 catalytic domain, HIS-NS4A₂₁32-PAG-NS3₃₋₁₈₁, was constructed by joining the NS4A peptide
GSVVIVGRIILS (NS4A amino acids 21-32) to the N-terminal domain of
NS3 protease (NS3 amino acids 3-181) via the linker PAG (SEQ ID NO:
47), and was cloned into the pET-28b+ vector as described above. Primers
were designed to generate a 613 base pair PCR fragment containing an
NdeI site followed by the NS4A peptide, the PAG linker, and amino
acids 3-181 of the NS3 catalytic domain at the 5' terminus and a stop
codon flanked by an EcoRI site at the 3' terminus. The 5' primer reads as
follows:

5' GATATACATATGGGTTCTGTTGTTATTGTTGGTAGAATTATTTT

ATCTCCTGCTGGTATCACGGCCTACTCCCAA 3' (SEQ ID NO: 48).

The 3' primer reads as follows:

5' CTCAGCGAATTCTCAAGACCGCATAGTAGTTTCCAT 3'

15 (SEQ ID NO: 49).

Plasmid-DNA (PMD-34-2), comprised of pET-22b+ vector encompassing the gene insert encoding HIS-NS3 (1-631) from 1b/BK strain was used as the template for PCR.

xii. HIS-NS4A₂₁₋₃₂-PAG-NS3₃₋₁₈₁/I17K (SEQ ID NO: 50)

A single amino acid mutant of HIS-NS4A₂₁₋₃₂-PAG-NS3₃₋₁₈₁ was constructed by creating a point mutation at position 17 of the NS3 domain of HIS-NS4A₂₁₋₃₂-PAG-NS3₃₋₁₈₁ construct as described above. Two oligonucleotide primers, each complementary to opposite strands of the template were generated which contains the point mutation which alters amino acid residue number 17 (isoleucine) to a lysine. The top strand primer was as follows:

5' CGGGGCCTACTIGGTTGCAAGATCACTAGCCTTACAGGC 3' (SEQ ID NO: 51).

The bottom strand reads as follows:

15

5' GCCTGTAAGGCTAGTGATCTTGCAACCAAGTAGGCCCCG3'

(SEQ ID NO: 52).

The template, HIS-NS4A $_{21-32}$ -PAG-NS3 $_{3-181}$ along with these two primers were utilized in a PCR reaction to generate this point mutation.

5 xiii. HIS-NS4A₂₁₋₃₂-GGS-NS3₃₋₁₈₁ (SEQ ID NO: 53)

An NS4A-tethered form of NS3 catalytic domain, HIS-NS4 $A_{21:32}$ -GGS-NS3 $_{3:181}$ was constructed by joining the NS4A peptide GSVVIVGRIILS (NS4A amino acids 21-32) to the N-terminal domain of NS3 protease (NS3 amino acids 3-181) via the linker GGS (SEQ ID NO: 54), and was cloned into the pET-28b+ vector as described above. Primers were designed to generate a 613 base pair PCR fragment containing an NdeI site followed by the NS4A peptide, the GGS linker, and amino acids 3-181 of the NS3 catalytic domain at the 5' terminus and a stop codon flanked by an EcoRI site at the 3' terminus. The 5' primer reads as follows:

5' GATATACATATGGGTTCTGTTGTTGTTGGTAGAATTATTTT

ATCTGGTGGTTCTATCACGGCCTACTCCCAA 3' (SEQ ID NO: 55).

The 3' primer reads as follows:

5' CTCAGCGAATTCTCAAGACCGCATAGTAGTTTCCAT 3'

20 (SEQ ID NO: 56).

Plasmid-DNA (PMD-34-2), comprised of pET-22b+ vector encompassing the gene insert encoding HIS-NS3 (1-631) from 1b/BK strain was used as the template for PCR.

xiv. HIS-NS4A₂₁₋₃₂-GGS-NS3₃₋₁₈₁/I17K (SEQ ID NO: 57)

A single amino acid mutant of HIS-NS4A₂₁₋₃₂-GGS-NS3₃₋₁₈₁ was constructed by creating a point mutation at position 17 of the NS3 domain of HIS-NS4A₂₁₋₃₂-GGS-NS3₃₋₁₈₁ construct as described above. Two oligonucleotide primers, each complementary to opposite strands of the template, were generated which contain the point mutation

15

20

which alters amino acid number 17 (isoleucine) to a lysine. The top strand primer was as follows:

5' CGGGGCCTACTTGGTTGCAAGATCACTAGCCTTACAGGC 3'

(SEQ ID NO: 58).

5 The bottom strand reads as follows:

5' GCCTGTAAGGCTAGTGATCTTGCAACCAAGTAGGCCCCG 3'

(SEQ ID NO: 59).

The template, HIS-NS4A $_{21-32}$ -GGS-NS3 $_{3-181}$, along with these two primers, was utilized in a PCR reaction to generate this point mutation.

EXAMPLE 2

NS3 Full-Length Constructs

i. HIS-NS31-631/I17K (SEQ ID NO: 60)

A single amino acid mutant of HIS-NS3₁₋₆₃₁ was formed by creating a point mutation at position 17 of NS3 protease using the Plasmid-DNA (PMD-34-2), comprised of pET-22b+ vector encompassing the gene insert, encoding HIS- NS3₁₋₆₃₁ from 1b/BK strain as described above. Two oligonucleotide internal primers, each complementary to opposite strands of the template, were generated which contain the point mutation which alters amino acid number 17 (isoleucine) to a lysine. The top strand primer was as follows:

5' CGGGGCCTACTTGGTTGCAAGATCACTAGCCTTACAGGC 3'

(SEQ ID NO: 61).

The bottom strand reads as follows:

25 5' GCCTGTAAGGCTAGTGATCTTGCAACCAAGTAGGCCCCG 3' (SEQ ID NO: 62).

15

20

The template, plasmid-DNA (PMD-34-2), comprised of pET-22b+ vector encompassing the gene insert, encoding HIS-NS3₁₋₆₃₁ from 1b/BK strain, along with these two primers was utilized in a PCR reaction to generate this point mutation.

5 ii. HIS-NS3₁₋₆₃₁/I18K (SEQ ID NO: 63)

A single amino acid mutant of HIS-NS3₁₋₆₃₁ was formed by creating a point mutation at position 18 of NS3 protease using the Plasmid-DNA (PMD-34-2), comprised of pET-22b+ vector encompassing the gene insert, encoding HIS-NS3₁₋₆₃₁ from 1b/BK strain as described above. Two oligonucleotide internal primers, each complementary to opposite strands of the template, were generated which contain the point mutation which alters amino acid number 18 (isoleucine) to a lysine. The top strand primer was as follows:

5' CGGGGCCTACTTGGTTGCATCAAGACTAGCCTTACAGGC 3'

(SEQ ID NO: 64).

The bottom strand reads as follows:

5' GCCTGTAAGGCTAGTCTTGATGCAACCAAGTAGGCCCCG 3'

(SEQ ID NO: 65).

The template, plasmid-DNA (PMD-34-2), comprised of pET-22b+ vector encompassing the gene insert, encoding HIS-NS3₁₋₆₃₁ from 1b/BK strain along with these two primers was utilized in a PCR reaction to generate this point mutation.

iii. HIS-NS31-631/S139A (SEQ ID NO: 66)

A single amino acid mutant of HIS-NS3₁₋₆₃₁ was formed by creating a point mutation at position 139 of the NS3 protease using the Plasmid-DNA (PMD-34-2), comprised of pET-22b+ vector encompassing the gene insert, encoding HIS-NS3₁₋₆₃₁ from 1b/BK strain as described above. Two oligonucleotide internal primers, each complementary to opposite strands of the template, were generated which contain the point mutation which altered amino acid number 139 (catalytic serine) to an alanine. The top strand primer was as follows:

25

5' CTCCTACTTGAAGGGCTCTGCTGGTGGTCCACTGCTCTGC 3'

(SEQ ID NO: 67).

The bottom strand reads as follows:

5' GCAGAGCAGTGGACCACCAGCAGAGCCCTTCAAGTAGGAG 3'

5 (SEQ ID NO: 68).

The template, plasmid-DNA (PMD-34-2), comprised of pET-22b+ vector encompassing the gene insert, encoding HIS-NS3 $_{1-631}$ from 1b/BK strain along with these two primers was utilized in a PCR reaction to generate this point mutation.

10 iv. HIS-NS3₁₋₆₃₁/I403S (SEQ ID NO: 69)

A single amino acid mutant of HIS-NS3₁₋₆₃₁ was formed by creating a point mutation at position 403 of the NS3 protease using the Plasmid-DNA (PMD-34-2), comprised of pET-22b+ vector encompassing the gene insert, encoding HIS-NS3₁₋₆₃₁ from 1b/BK strain as described above. Two oligonucleotide internal primers, each complementary to opposite strands of the template, were generated which contain the point mutation which alters amino acid number 403 (isoleucine) to a serine. The top strand primer was as follows:

5' GTCCGTCATACCAACTTCCGGAGACGTCGTTGTCG 3'

20 (SEQ ID NO: 70).

The bottom strand reads as follows:

5' CGACAACGACGTCTCCGGAAGTTGGTATGACGGAC 3'

(SEQ ID NO: 71).

The template, plasmid-DNA (PMD-34-2), comprised of pET-22b+ vector encompassing the gene insert, encoding HIS-NS3 $_{1-631}$ from 1b/BK strain along with these two primers was utilized in a PCR reaction to generate this point mutation.

v. HIS-NS3₁₋₆₃₁/NdeI (SEQ ID NO. 72)

A silent mutant of HIS-NS3₁₋₆₃₁ was formed to eliminate the internal NdeI restriction site within NS3 protease using the Plasmid-DNA (PMD-34-2), comprised of pET-22b+ vector encompassing the gene insert, encoding HIS-NS3₁₋₆₃₁ from 1b/BK strain as described above. Two oligonucleotide internal primers, each complementary to opposite strands of the template, were generated which contain point mutations which alters the codons on the reading strand of alanine 217 from GCA to GCC and tyrosine 218 from TAT to TAC. The top strand primer was as follows:

5' ACTAAAGTGCCGGCTGCCTACGCAGCCCAAGGG 3'

(SEQ ID NO: 73).

The bottom strand reads as follows:

5' CCCTTGGGCTGCGTAGGCAGCCGGCACTTTAGT 3'

(SEQ ID NO: 74).

15 The template, plasmid-DNA (PMD-34-2), comprised of pET-22b+ vector encompassing the gene insert encoding HIS-NS3₁₋₆₃₁ from 1b/BK strain, along with these two primers, was utilized in a PCR reaction to generate this point mutation.

vi. HIS-NS4A21-32-GSGS-NS33-631 (SEQ ID NO: 4)

An NS4A-tethered form of the NS3 full-length domain, HIS-20 NS4A₂₁₋₃₂-GSGS-NS3₃₋₆₃₁ was constructed via a cut and paste strategy as described above. Briefly, a 270 bp fragment was generated by restricting HIS-NS4A21-32-GSGS-NS33-181 with XbaI/BspMI; This fragment encompassed sequences encoding a histidine tag followed by a thrombin 25 site, the NS4A peptide, GSVVIVGRIILS (NS4A amino acids 21-32), the linker GSGS (SEQ ID NO: 21) and NS3 amino acids 3-48. A second 7111 fragment (7111 bp) was generated by restricting Plasmid-DNA (PMD-34-2), comprised of pET-22b+ vector encompassing the gene insert. encoding HIS-NS3 (1-631) from 1b/BK strain with XbaI/BspmI resulting 30 in a fragment encompassing the pET 22b+ vector backbone in addition to amino acids 49-631. These two fragments were then ligated together with T4 DNA ligase to form HIS-NS4A21-32-GSGS-NS33-631.

20

25

vii. HIS-NS4A₂₁₋₃₂-GSGS-NS3₃₋₆₃₁/I17K (SEO ID NO: 12)

A single amino acid mutant of HIS-NS4A $_{21.32}$ -GSGS-NS3 $_{3.631}$ was constructed by creating a point mutation at position 17 of the NS3 domain of HIS-NS4A $_{21.32}$ -GSGS-NS3 $_{3.631}$ construct as described above. Two oligonucleotide primers, each complementary to opposite strands of the template, were generated which contain the point mutation which alters amino acid number 17 (isoleucine) to a lysine. The top strand primer was as follows:

10 5' CGGGGCCTACTTGGTTGCAAGATCACTAGCCTTACAGGC 3'

(SEQ ID NO: 75).

The bottom strand read as follows:

5' GCCTGTAAGGCTAGTGATCTTGCAACCAAGTAGGCCCCG 3'

(SEQ ID NO: 76).

15 The template, HIS-NS4A₂₁₋₃₂-GSGS-NS3₃₋₆₃₁ along with these two primers was utilized in a PCR reaction to generate this point mutation.

viii. HIS-NS4A₂₁₋₃₂-GSGS-NS3₃₋₆₃₁/I18K (SEO ID NO: 13)

A single amino acid mutant of HIS-NS4A $_{21.32}$ -GSGS-NS3 $_{3.631}$ was constructed by creating a point mutation at position 18 of the NS3 domain of HIS-NS4A $_{21.32}$ -GSGS-NS3 $_{3.631}$ construct as described above. Two oligonucleotide primers, each complementary to opposite strands of the template were generated which contained the point mutation which alters amino acid number 18 (isoleucine) to a lysine. The top strand primer was as follows:

5' CGGGGCCTACTTGGTTGCATCAAGACTAGCCTTACAGGC 3'

(SEQ ID NO: 77).

The bottom strand reads as follows:

5' GCCTGTAAGGCTAGTCTTGATGCAACCAAGTAGGCCCCG 3'

(SEQ ID NO: 78).

The template, HIS-NS4A $_{21-32}$ -GSGS-NS3 $_{3-631}$, along with these two primers was utilized in a PCR reaction to generate this point mutation.

HIS-NS4A₂₁₋₃₂-GSGS-NS3₃₋₆₃₁/I17K, I18K (SEO ID: 14)

A double amino acid mutant of HIS-NS4A₂₁₋₃₂-GSGS-NS3₃₋₆₃₁ was constructed by creating 2 point mutations at positions 17 and 18 of the NS3 domain of the HIS-NS4A₂₁₋₃₂-GSGS-NS3₃₋₆₃₁ construct simultaneously as described above. Two oligonucleotide primers, each complementary to opposite strands of the template, were generated which contain the point mutations which alter amino acid numbers 17 (isoleucine) and 18 (isoleucine) to lysines. The top strand primer was as follows:

5' CGGGGCCTACTTGGTTGCAAGAAGACTAGCCTTACAGGC 3'

(SEQ ID NO: 79).

15 The bottom strand read as follows:

5' GCCTGTAAGGCTAGTCTTCTTGCAACCAAGTAGGCCCCG 3'

(SEO ID NO: 80).

The template, HIS-NS4A₂₁₋₃₂-GSGS-NS3₃₋₆₃₁, along with these two primers, was utilized in a PCR reaction to generate this point mutation.

20 x. HIS-NS4A₂₁₋₃₂-GSGS-NS3₃₋₆₃₁/S139A (SEQ ID NO: 15)

An NS4A-tethered form of NS3 full-length domain, HIS-NS4A₂₁. 32-GSGS-NS3₃₋₆₃₁/S139A, was constructed via a cut and paste strategy as described above. Briefly, a 290 bp fragment was generated by restricting HIS-NS4A₂₁₋₃₂-GSGS-NS3₃₋₁₈₁ with XbaI/BspMI; this fragment encompass sequence encoding a histidine tag, a thrombin site, amino acids 21-32 of the the NS4A peptide, the linker GSGS (SEQ ID NO. 21) and NS3 amino acids 3-48. A second 7111 fragment (7111 bp) was generated by restricting HIS-NS3₁₋₆₃₁/S139A construct with XbaI/BspmI resulting in a fragment encompassing the pET 22b+ vector backbone in addition to amino acids

10

15

49-631. These two fragments were then ligated together with T4 DNA ligase to form HIS-NS4A₂₁₋₃₂-GSGS-NS3₂₋₆₃₁/S139A.

xi. HIS-NS4A₂₁₋₃₂-GSGS-NS3₃₋₆₃₁/S139A, I17K (SEQ ID NO: 16)

A single amino acid mutant of HIS-NS4A $_{21-32}$ -GSGS-NS3 $_{3-631}$ /S139A was constructed by creating a point mutation at position 17 of the NS3 domain of the HIS-NS4A $_{21-32}$ -GSGS-NS3 $_{3-631}$ /S139A construct as described above. Two oligonucleotide primers, each complementary to opposite strands of the template, were generated which contain the point mutation which alters amino acid number 17 (isoleucine) to a lysine. The top strand primer was as follows:

5' CGGGGCCTACTTGGTTGCAAGATCACTAGCCTTACAGGC 3'

(SEQ ID NO: 81).

The bottom strand is as follows:

5'GCCTGTAAGGCTAGTGATCTTGCAACCAAGTAGGCCCCG 3'

(SEO ID NO: 82).

The template HIS-NS4A $_{21-32}$ -GSGS-NS3 $_{3-631}$ /S139A, along with these two primers, was utilized in a PCR reaction to generate this point mutation.

xii. HIS-NS4A₂₁₋₃₂-GSGS-NS3₃₋₆₃₁/S139A, I18K (SEQ ID NO: 17)

A single amino acid mutant of HIS-NS4A₂₁₋₃₂-GSGS-NS3₃₋₆₃₁/S139A was constructed by creating a point mutation at position 18 of the NS3 domain of the HIS-NS4A₂₁₋₃₂-GSGS-NS3₃₋₆₃₁/S139A construct as described above. Two oligonucleotide primers, each complementary to opposite strands of the template, were generated which contain the point mutation which alters amino acid number 18 (isoleucine) to a lysine. The top strand primer was as follows:

5' CGGGGCCTACTTGGTTGCATCAAGACTAGCCTTACAGGC 3' (SEQ ID NO: 83).

The bottom strand read as follows:

5' GCCTGTAAGGCTAGTCTTGATGCAACCAAGTAGGCCCCG 3'

(SEQ ID NO: 84).

The template HIS-NS4A₂₁₋₃₂-GSGS-NS3₃₋₆₃₁/S139A, along with these two primers, was utilized in a PCR reaction to generate this point mutation.

xiii. HIS-NS4A₂₁₋₃₂-GSGS-NS3₃₋₆₃₁/S139A, I17K, I18K (SEQ ID NO: 18)

A single amino acid mutant of HIS-NS4A $_{21-32}$ -GSGS-NS3 $_{3-631}$ /S139A, I17K was constructed by creating a point mutation at position 18 of the NS3 domain of the HIS-NS4A $_{21-32}$ -GSGS-NS3 $_{3-631}$ /S139A, I17K construct as described above. Two oligonucleotide primers, each complementary to opposite strands of the template, were generated which contain the point mutation which alters amino acid number 18 (isoleucine) to an lysine. The top strand primer was as follows:

5' CGGGGCCTACTTGGTTGCAAGAAGACTAGCCTTACAGGC 3'

15 (SEQ ID NO: 85).

The bottom strand reads as follows:

5' GCCTGTAAGGCTAGTCTTCTTGCAACCAAGTAGGCCCCG 3'

(SEQ ID NO: 86).

The template HIS-NS4A₂₁₋₃₂-GSGS-NS3₃₋₆₃₁/S139A,I17K, along with 20 these two primers was utilized in a PCR reaction to generate this point mutation.

xiv. HIS-NS4A₁₅₋₃₂-GSGS-NS3₃₋₆₃₁ (SEQ ID NO: 19)

A NS4A-tethered form of NS3 full-length domain, HIS-NS4A₂₁₋₃₂-GSGS-NS3₃₋₆₃₁ was constructed by joining the amino acids 15-32 of NS4A peptide to the N-terminal end of the NS3 protease (NS3 amino acids 3-631) via the linker GSGS, and was cloned into the pET-28b+ vector as described above with the following modification. Primers were designed to generate a PCR fragment containing an NdeI site followed by the NS4A peptide, the GSGS linker (SEQ ID NO: 21), and amino acids 3-631

20

25

30

of the NS3 catalytic domain at the 5' terminus and a stop codon flanked by an EcoRI site at the 3' terminus. The 5' primer sequence was as follows:

5'GATATACATATGGCTTACTCTCTGACTACGGGTTCTGTTGTTATT

5 GTTGGTAGAATTATTTTATCTGGTAGTGGTAGTATCACGGCCTACTCCCAA 3'

(SEO ID NO: 87).

The 3' primer sequence was as follows: 5' GTGGTGGTGCTCGAGGCTGCCGCGCGCA

CCAGCGTAACGACCTCCAGGTC 3' (SEO ID NO: 88).

10 The template used was HIS-NS4A₂₁₋₃₂-GSGS-NS3₃₋₆₃₁. The resulting PCR fragment was 1974 bases. Vent DNA polymerase was employed and a final concentration of 200 µM dNTPS was used. The PCR conditions were as follows: 95 $^{\circ}\text{C}$ for 45 seconds (1 cycle); 95 $^{\circ}\text{C}$ for 30 seconds, 55 °C for 1 minute, 72 °C for 2 minutes (25 cycles). The product was purified with QIAquick PCR kit (Qiagen). This PCR product, along with the 6.6 kb vector backbone (HIS-NS4A21-32-GSGS-NS33-631), were double digested with NdeI and BamHI. The digested fragments of 1.43 and 6.6 Kbp respectively were run on agarose gel, excised, and column purified with QIAquick gel extraction kit (Qiagen). They were quantitated and then ligated together with T4 DNA ligase.

xv.HIS-NS4A₁₅₋₃₂-GSGS-NS3₃₋₆₃₁/S139A (SEQ ID NO: 20)

An NS4A-tethered form of NS3 full-length domain, HIS-NS4A21-32-GSGS-NS33-631/S139A was constructed by joining amino acids 15-32 of the NS4A peptide to the N-terminal end of the NS3 protease (NS3 amino acids 3-631) via the linker GSGS (SEQ ID NO: 21), and was cloned into the pET-28b+ vector as described above with the following modification. Primers were designed to generate a PCR fragment containing an NdeI site followed by the NS4A peptide, the GSGS linker (SEQ ID NO: 21), and amino acids 3-631 of the NS3 catalytic domain at the 5' terminus and a stop codon flanked by an EcoRI site at the 3' terminus. The 5' primer sequence was as follows:

15

25

30

5'GATATACATATGGCTTACTCTCTGACTACGGGTTCTGTTGTTATT

GTTGGTAGAATTATTTTATCTGGTAGTGGTAGTATCACGGCCTACTCCCAA 3'

(SEQ ID NO: 89).

The 3' primer reads as follows:

5 5' TGGTGGTGCTCGAGGCTGCCGCGCGCACCAGCGTAACGACCT

CCAGGTC 3' (SEQ ID NO: 90).

The template used was HIS-NS4A $_{21:32}$ -GSGS-NS3 $_{3:631}$ /S139A. The resulting PCR fragment was 1974 bases. Vent DNA polymerase was employed and a final concentration of 200 μ M dNTPS was used. The PCR conditions were as follows: 95 °C for 45 seconds (1 cycle); 95 °C for 30 seconds, 55 °C for 1 minute, 72 °C for 2 minutes (25 cycles). The product was purified with QIAquick PCR kit (Qiagen). This PCR product along with the 6.6 kb vector backbone (HIS-NS4A $_{21:32}$ -GSGS-NS3 $_{3:631}$) were double digested with NdeI and BamHI. The digested fragments of 1.43 and 6.6 Kbp respectively were run on agarose gel, excised, and column purified with QIAquick gel extraction kit (Qiagen). They were quantitated and then ligated together with T4 DNA ligase.

EXAMPLE 3

20 <u>Expression and Purification of HCV NS4A-NS3 Complexes</u>

A. Small Scale Expression Studies

All constructed plasmids were transformed into DH5 α cells for production of large amount of plasmid-DNA. The purified plasmid-DNA was transformed into BL21(DE3) cells for expression studies. The cells were grown in Terrific Broth in baffled flasks at 37°C to an OD of 1.0 and the temperature was lowered to 23°C. The cultures were induced with 0.4 mM IPTG and were harvested 3 hours after induction. Cells were sonicated for 1 min in 50 mM HEPES, pH 7.5, 20% glycerol, 0.1% β OC, 0.3 M NaCl, 10 mM β ME and spun at 13,000 rpm for 10 min. The supernatants were analyzed on 10% Novex SDS-PAGE.

15

20

25

30

35

B. Large-Scale Expression And Purification Of NS4A-Tethered Forms Of HCV NS33-181 Protease

 $E.\ coli,$ BL21(DE3) cells harboring either plasmid pET-22b or pET-28b encoding various native, single, or multiple mutants of NS4A-tethered forms of NS31-181 were grown at 37°C in Terrific Broth supplemented with either 100 ug/ml of ampicillin (for pET-22b) or 25 ug/ml kanamycin (for pET28-b) in 10-liter fermentor. When the cell density reaches an OD of 2-3, the temperature was lowered to 23°C within 5 minutes and cells were induced with 0.4 mM IPTG. Cells were harvested 3 hours after induction and frozen at -20 °C prior to purification.

Cell pellets were resuspended in 600 ml of lysis buffer containing 50 mM HEPES, pH 7.4, 10% glycerol, 0.3 M NaCl, 0.1% βOG, 2 mM βME (buffer A), homogenized using a cell homogenizer (Omni Mixer ES) for 2 min and the cells were disrupted by two passes through a Microfluidizer (Microfluidics Model #M-110F) at 10,000 p.s.i. The lysate was centrifuged at 85,000 x g for 45 min. The supernatant was filtered through 0.8 micron filter units (Nalgene) and applied at 40 ml/min to a 11-ml Ni-imidodiacetate (POROS 20 MC resin) column in the presence of 20 mM immidazole on BIOCAD (Perseptive Biosystems). column was washed with 10 column volumes of buffer A, followed by 15 column volume of buffer A containing 1.0 M NaCl and 20 mM imidazole (buffer B). The bound protease was eluted with the elution buffer (buffer B containing 250 mM imidazole). The eluted fractions containing the protease were pooled and dialyzed versus 16 liters of 50 mM HEPES, pH 7.4, 10% glycerol, 1 M NaCl, 10 mM ßME in order to remove the imidazole and the detergent.

When the removal of the N-terminal histidine tag was required, human thrombin (Enzyme Research) was added to the eluted, pooled fractions at a thrombin:protease ratio of 8 units per mg of protease and thrombin cleavage was allowed to proceed during the dialysis step for 18 hours. The dialyzed, thrombin-cleaved protease was applied to 3 sephacryl-100 sizing column (26 × 60cm, Pharmacia) in series, equilibrated in of 50 mM HEPES, pH 7.4, 10% glycerol, 1 M NaCl, 10 m M bME at 0.5 ml/min. Fractions containing purified protease at above

15

20

25

35

>95% homogeneity as judged by SDS-PAGE were pooled and flashfrozen at -80 °C

C. Large-Scale Expression And Purification Of NS4A-Tethered Forms Of HCV NS33-631 Protease

E. coli, BL21(DE3) cells harboring either plasmid pET-22b or pET-28b encoding various native, single, or multiple mutants of NS4Atethered forms of NS31-181 were grown at 37°C in Terrific Broth supplemented with either 100 µg/ml of ampicillin (for pET-22b) or 25 μg/ml kanamycin (for pET28-b) in 10-liter fermentor. When the cell density reaches an OD of 2-3, the temperature was lowered to 23°C within 5 minutes and cells were induced with 0.4 mM IPTG. Cells were harvested 3 hours after induction and frozen at -20 °C prior to purification.

Cell pellets were resuspended in 600 ml of lysis buffer containing 50~mM HEPES, pH 7.4, 10% glycerol, 0.3 M NaCl, 0.1% βOG , 2 mM βME (buffer A), homogenized using a cell homogenizer (Omni Mixer ES) for 2 min and the cells were disrupted by two passes through a Microfluidizer (Microfluidics Model #M-110F) at 10,000 p.s.i. The lysate was centrifuged at 85,000 x g for 45 min. The supernatant was filtered through 0.8 micron filter units (Nalgene) and applied at 40 ml/min to a 11-ml Ni-imidodiacetate (POROS 20 MC resin) column in the presence of 20 mM immidazole on BIOCAD (Perseptive Biosystems). column was washed with 10 column volumes of buffer A, followed by 15 column volume of buffer A containing 1.0 M NaCl and 20 mM imidazole (buffer B). The bound protease was eluted with the elution buffer (buffer B containing 250 mM imidazole). The eluted fractions containing the protease were pooled and dialyzed versus 16 liters of 50 mM HEPES, pH 7.4, 10% glycerol, 1 M NaCl, 10 mM ßME in order to remove the imidazole and the detergent.

30 When the removal of the N-terminal histidine tag was required, human thrombin (Enzyme Research) was added to the eluted, pooled fractions at a thrombin:protease ratio of 8 units per mg of protease and thrombin cleavage was allowed to proceed during the dialysis step for 18 The dialyzed, thrombin-cleaved protease was applied to 3 sephacryl-100 sizing column (26 x 60cm, Pharmacia) in series,

equilibrated in of 50 mM HEPES, pH 7.4, 10% glycerol, 1 M NaCl, 10 m M β ME at 0.5 ml/min. Fractions containing purified protease at above >95% homogeneity as judged by SDS-PAGE were pooled and flash-frozen at -80 °C.

5

10

15

20

30

EXAMPLE 4

Molecular Weight Determination Of Various NS3 Protease Forms By Size Exclusion Chromatography

Two hundred μl of various purified proteins were applied to a calibrated Superdex-75 HR (1cm x 30 cm) FPLC column equilibrated with 25 mM HEPES, pH 7.4, 1M NaCl and 10% glycerol and 10 mM β ME at 0.5 ml/min. The column was precalibrated using Pharmacia standard calibration proteins (BSA: 67 KDa; Ovalbumin: 43 KDa; Chymotrypsinogen A: 31 KDa; Ribonuclease A: 13.7 KDa). Protein elution was monitored at 280 nm.

The following covalent NS4A-NS3 complexes described above were characterized by the above method:

HIS-NS4A₂₁₋₃₂-GSGS-NS3₃₋₁₈₁/117K HIS-NS4A₂₁₋₃₂-GSGS-NS3₃₋₁₈₁/117K HIS-NS4A₂₁₋₃₂-GSGS-NS3₃₋₁₈₁/118K HIS-NS4A₂₁₋₃₂-GSGS-NS3₃₋₁₈₁/S139A HIS-NS4A₂₁₋₃₂-GSGS-NS3₃₋₁₈₁/S139A, 117K HIS-NS4A₂₁₋₃₂-GSGS-NS3₃₋₁₈₁/S139A, 118K

25 HIS-NS4A₂₁₋₃₂-PAG-NS3₃₋₁₈₁/I17K

HIS-NS4A₂₁₋₃₂-GSGS-NS3₃₋₆₃₁/I17K HIS-NS4A₂₁₋₃₂-GSGS-NS3₃₋₆₃₁/I17K HIS-NS4A₂₁₋₃₂-GSGS-NS3₃₋₆₃₁/I18K HIS-NS4A₂₁₋₃₂-GSGS-NS3₃₋₆₃₁/S139A HIS-NS4A₂₁₋₃₂-GSGS-NS3₂₋₆₃₁/S139A, I17K

HIS-NS4A₂₁₋₃₂-PAGG-NS3₃₋₁₈₁/I17K

HIS-NS4A₂₁₋₃₂-GSGS-NS3₃₋₆₃₁/S139A, I18K

15

20

25

Of those constructs characterized, all covalent NS4A-NS3 complexes containing a three amino acid linker resulted in aggregated forms, as judged by size exclusion chromatography. NS4A-tethered forms in which a point mutation at position 17 or 18 had not been introduced also resulted in aggregated forms, although they exhibited activity identical to that of the monodispersed forms of the protease.

Covalent NS4A-NS3 complexes which contained a four amino acid linker <u>and</u> a point mutation at position 17 and/or 18 resulted in active, monodispersed proteins with apparent molecular weights smaller than predicted as determined by size exclusion chromatography.

EXAMPLE 5

Determination of Proteolytic Activity

Following expression and purification, newly engineered recombinant species were assayed for proteolytic activity utilizing a 1D-HPLC (reverse-phase chromatography) technique. conducted using the 5A/5B (P8P8') substrate DTEDVVCC*SMSYTWTG-K (SEQ ID NO: 25) in 25 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.5 m M EDTA, 10 mM DTT, 10% glycerol, and 0.05% lauryl maltoside. Concentration of all proteins were determined by BIORAD dye method). The catalytic domain His-NS3₁₋₁₈₁ (batch # 51072-92E) was preincubated at a concentration of 250 nM in the presence of 20 μM 4A peptide (KKGSVVIVGRIVLSGKPAIIPKK) for 15 minutes at 4°C. This mixture was then diluted into the reaction volume at a final concentration of 8 μM 4A peptide and 100 nM catalytic domain. Reactions were incubated at room temperature for 60 minutes and were quenched with an equal volume of 10% phosphoric acid. Following injection, cleavage products were monitored under a linear 0-80% acetonitrile gradient in 0.1% TFA. The product P1'P8'K peak areas were automatically converted to product quantity in nanomoles by a standard curve.

The various covalent NS4A-NS3 complexes whose proteolytic efficiency has been determined according to the above method, and the results of each determination, are shown in Table 1.

Table 1.

Construct	k _{cat} (min ⁻¹)	K _m (μM)	$k_{cat}/K_{m} (M^{-1} s^{-1})$			
NS3 ₁₋₆₉₁ -NS4A ₁₋₅₄	10 ± 2	20 ± 2	(8 ± 2) X 10 ³			
His-NS3 ₁₋₁₈₁ + NS4A Peptide	3±1	80 ± 20	$(0.5 \pm 0.2) \times 10^3$			
His-NS4A ₂₁₋₃₂ -GSGS-NS3 ₃₋₁₈₁	9±2	19 ± 3	(8 ± 2) X 10 ³			
His-NS4A ₂₁₋₃₂ -GSGS-NS3 ₃₋₁₈₁ /I17K	16±3	20 ± 2	(14 ± 2) X 10 ³			
His-NS4A ₂₁₋₃₂ -GSGS-NS3 ₃₋₁₈₁ /I18K	10 ± 2	22 ± 2	(8±2) X 10 ³			

 a [E] = 0.25 μ M, [NS4A Peptide] = 10 μ M

As can be seen from the forgoing results, all covalent NS4A-NS3 complexes were shown to have an equivalent catalytic efficiency to that of full-length NS3 $_{1-631}$ -NS4A $_{1-54}$. In contrast, the noncovalent complex of NS3 $_{1-181}$ with the NS4A peptide (0.1:8 μ M), KK-(NS4A $_{21-39}$)-KK, had an catalytic activity which is 8 fold lower than the full-length NS3 $_{1-631}$ -NS4A $_{1-54}$.

Example 6

High Throughput Screening Assays Using Covalent NS4A-NS3 Complexes

15

10

5

The claimed covalent NS4A-NS3 complexes are useful in screening methods for identifying NS3 protease inhibitors. One such method in which the claimed covalent complexes can be used is illustrated below.

20

Surface Plasmon Resonance Assay

The present example illustrates a method for determining if a compound can be useful as an HCV protease inhibitor using the surface plasmon resonance assay. Figures 5A and 5B schematically depict the technique.

BIAcore[™] is a processing unit for Biospecific Interaction Analysis.

The processing unit integrates an optical detection system with an autosampler and a microfluidic system. BIAcore[™] uses the optical

phenomena of surface plasmon resonance to monitor interaction between biomolecules.

SPR is a resonance phenomenon between incoming photons and electrons on the surface of thin metal film. Resonance occurs at a sharply defined angle of incident light. At this angle, called the resonance angle, energy is transferred to the electrons in the metal film, resulting in a decreased intensity of the reflected light. SPR response depends on a change in refractive index in the close vicinity of the sensor chip surface, and is proportional to the mass of analyte bound to the surface. The BIAcore™ continuously measures the resonance angle by a relative scale of resonance units (RU) and displays it as an SPR signal in a sensorgram, where RU are plotted as a function of time.

BIAcore we uses continuous flow technology. One interactant is immobilized irreversibly on the sensor chip, comprising a noncrosslinked carboxymethylated dextran providing a hydrophilic environment for bimolecular interaction. Solution containing the other interactant flows continuously over the sensor chip surface. As molecules from the solution bind to the immobilized ligand, the resonance angle changes resulting in a signal registered by the instrument.

In this methodology, the enzymatic reactions are carried out outside of the $BIAcore^{TM}$, in reaction tubes or 96-well tissue culture plates, as it is conventionally done for any of the other available high throughput assays. The SPR is only used as a detection means for determination of the amount of an intact substrate remaining in a solution after the reaction is quenched.

30

35

10

15

20

25

In order to measure the amount of the intact substrate prior to the addition of enzyme, a means of capturing the substrate onto the sensor chip had to be established. In addition, to satisfy the requirement for a high throughput assay on the $BIAcore^{TM}$, the substrate needed to be removed from the surface after completion of analysis, so that the same surface can be used for subsequent reactions. To accomplish these two requirements, a phosphotyrosine is synthetically attached to one end of the substrate. The phosphotyrosine was chosen due to the commercial

10

15

20

25

30

35

availability of an anti-phosphotyrosine monoclonal antibody. The antibody is covalently attached to the sensor chip by standard amine coupling chemistry. The anti-phosphotyrosine antibody, bound permanently to the chip, is used to capture the phosphotyrosine in a reversible manner. The antibody-phosphotyrosine interaction is ultimately used to capture and release the attached peptide substrate. After completion of analysis, the surface can be regenerated using various reagents such as 2 M MgCl2.

When an intact peptide substrate is introduced onto the antibody surface, a large mass is detected by the instrument. To follow the extent of peptide cleavage, a mixture of peptide substrate and enzyme is incubated for the desired time and then quenched. Introduction of this mixture, containing both cleaved peptide and intact peptide, to a regenerated antibody surface results in detection by the instrument of a lower mass than that detected for the sample containing only intact peptide. The difference in the two values is then used to calculate the exact amount of intact peptide remaining after cleavage by the enzyme.

Although the reduction in mass can be directly followed with many large substrates, due to the small mass of a typical synthetic peptide substrate (10-20 amino acids, 1-3 Daltons), the mass difference, and thus the signal difference between the intact and cleaved peptide, is very small within the signal to noise ratio of the instrument. To circumvent this low sensitivity, a biotin can be attached at the N-terminus of the peptide. Streptavidin can then be added, thus tagging the peptide. When the tagged peptide is introduced onto the antibody surface of the chip, the signal will be higher. The signal resulting from introduction of a cleaved peptide which lacks the N-terminal half, (and thus the streptavidin), will be much lower.

To carry out this method, an HCV protease 5A-5B peptide substrate, (such as 5A/5B substrate DTEDVVACSMSYTWYG-K (SEQ ID NO: 91)) is synthesized with an additional phosphotyrosine at the C-terminus and a biotin at the N-terminus. The biotin is then tagged with streptavidin. An anti-phosphotyrosine monoclonal antibody, 4G10 (Upstate Biotechnology Inc., Lake Placid, New York) is coupled to the sensor chip. In the absence of an active, uninhibited HCV protease,

20

25

introduction of the intact phosphotyrosine peptide results in a large signal (large mass unit/large signal) through its interaction with the anti-phosphotyrosine monoclonal antibody (Mab).

5 The protease-catalyzed hydrolysis of the phosphotyrosine-biotinylated peptide is carried out in a 96 well plate. The reaction is stopped with an equal volume of mercuribenzoate. The cleaved peptide which lacks the tagged streptavidin (less mass) results in the loss of response units (lower signal).

Using this method, numerous compounds can be tested for their inhibitory activity since the antibody surface can be regenerated repetitively with 2 M MgCl₂.

15 Procedure for Coupling Anti-phosphotyrosine Mab to the Sensor Chip

The anti-phosphotyrosine Mab is coupled to the carboxymethylated dextran surface of a sensor chip in the following manner. The flow rate used throughout the coupling procedure is 5 $\mu l/min$. The surface is first activated with a 35 μl injection of NHS/EDC (N-hydroxysuccinimide/N-dimethyllaminopropyl-N'-ethylcarbodiimide-HCl). This is followed by a 40 ml injection of Mab 4G10 at 50 $\mu g/ml$ in 10 mM sodium acetate buffer, pH=4.0. Any remaining activated esters are then blocked by the injection of 35 μl of 1 M ethanolamine. These conditions result in the immobilization of approximately 7,500 response units (420 μM) of antibody.

Binding of Peptide and Regeneration of Mab 4G10 Surface

30 The flow rate used throughout the BIAcore analysis run is 5 μ l/min. A 4 μ l injection containing streptavidin-tagged peptide (peptide concentration at 2 μ M, streptavidin binding sites concentration at 9 μ M) is carried out. The amount of streptavidin-tagged peptide bound to the antibody surface (in response units) is measured 30 seconds after the 35 injection is complete.

Regeneration of the Mab 4G10 surface is achieved using a 4 μl pulse of 2 M MgCl $_2$ after each peptide injection. Surfaces regenerated up to 500 times still showed 100% binding of tagged peptide.

5 <u>Determination of the Optimal Concentration of Peptide and</u> Streptavidin

To determine the optimal peptide concentration, a standard curve was generated using various amounts of peptide (0-10 μ M) in the presence of excess streptavidin. A value in the linear range, 2 μ M, was chosen for standard assay conditions.

The amount of streptavidin required to completely tag the peptide is determined using a peptide concentration of 2.5 μM and titrating the amount of streptavidin (μM of binding sites). All the peptides were shown to be completely tagged when streptavidin concentrations greater than 3 μM (approximately equimolar to the peptide concentration) were used. A streptavidin concentration of 9 μM (a 4.5 fold excess) was chosen for standard assay conditions.

Application of Described Methodology to Covalent HCV NS4A-NS3 Complexes

The HCV protease 5A/5B peptide substrate, (such as 5A/5B substrate DTEDVVACSMSYTWYG-K (SEQ ID NO: 91)), with a phophotyrosine synthetically attached to the C-terminus and a biotin attached at the N-terminus, is synthesized. Anti-phosphotyrosine monoclonal antibody, 4G10 is coupled to the sensor chip.

In the absence of active, uninhibited covalent HCV NS4A-NS3 complex, the introduction of the intact streptavidin-tagged biotinylated phosphotyrosine peptide to the sensor chip results in a large signal (large mass unit/large response units) through its interaction with the anti-phosphotyrosine monoclonal antibody.

The protease-catalyzed hydrolysis of the phosphotyrosinebiotinylated peptide is carried out with and without a suspected inhibitor

30

35

25

10

15

in a 96 well plate. The reaction is stopped with an equal volume of the quenching buffer containing mercuribenzoate. Streptavidin is then added to tag the peptide. The cleaved peptide, which lacks the streptavidin (less mass), results in the loss of response units.

5

15

20

25

30

35

Using this assay, numerous compounds can be tested for their inhibitory activity since the antibody surface can be regenerated repetitively with 2 M MgCl $_2$.

10 Standard Operating Procedure for BIAcore-based HCV Assay

Reactions are prepared in a 96-well tissue culture plate using the Reaction Buffer (50 mM HEPES, pH 7.4, 20 % glycerol, 150 mM NaCl, 1mM EDTA, 0.1% Tween-20,1 mM DTT) as diluent. The final reaction volume is 100 $\mu l. \,$ Sample with the peptide alone (Biotin-DTEDVVAC SMSYTWTGKpY) is prepared by addition of 10 μ l of peptide stock at 100 μM (prepared in the reaction buffer) to 90 μl of reaction buffer, so that the final concentration of peptide is 10 µM. Samples comprised of peptide and the covalent NS4A-NS3 complexes are prepared by addition of 10 μl of peptide stock at 100 μM and 10 μl of covalent NS4A-NS3 stock at 0.17 mg/ml (both prepared in the reaction buffer) to 80 µl of reaction buffer, so that the final concentration of peptide and the enzyme is 10 and 0.1 µM respectively. The reaction is held at 30°C for the specified time and then quenched. Quenching is achieved by transferring a $20-\mu l$ aliquot of the reaction mixture to a new tissue culture plate containing an equal volume of PMB Quenching Buffer (50 mM HEPES, pH 7.8, 150 mM NaCl, 5 mM P-Hydroxymercuribenzoic Acid, and 13 mM EDTA).

To prepare the quenched reaction mixture for injection onto the sensor surface, 30 μ l PMB BIAcore Buffer (50 mM HEPES, pH 7.4, 1 M NaCl) and 30 μ l of streptavidin at 0.5 mg/ml in water is added to the 40 μ l of the quenched reaction mixture to a final volume of 100 μ l. In this step, all the peptides are tagged with streptavidin prior to the injection of samples. Finally, 4 μ l of this sample is injected over the antiphosphotyrosine surface for determination of the intact versus cleaved peptide. The final concentration of peptide and the streptavidin in the BIAcore sample is 2 and 9 μ M, respectively.

Experimental Conditions:

Substrate: Biotin-DTEDVVAC SMSYTWTGK-pY (SEQ

ID NO: 91) in Reaction buffer without DTT

Concentration:

170 μM (Crude peptide, based on weight)

Enzyme:

10 μl of concentrated His-NS4A21-32-GSGS-

NS33-181 at 0.17 mg/ml

10

5

Reaction volume:

100 µl

Reaction buffer:

50 mM HEPES, pH 7.8

20 % glycerol 150 mM NaCl

15

1mM EDTA 1mM DTT 0.1% Tween-20

20 <u>Temp</u>:

30° C

Quench with:

p-hydroxymercuribenzoate

25

30

35

EXAMPLE 7

Determination of Nucleic Acid Unwinding Activity

The newly engineered single-chain recombinant His-NS4A₂₁₋₃₂-GSGS-NS3₃₋₆₃₁ (SEQ ID NO: 4) was assayed for nucleic acid unwinding activity using a scintillation proximity assay (SPA, Amersham Life Science Inc., Arlington Height, IL). The unwinding activity present in this covalent His-NS4A₂₁₋₃₂-GSGS-NS3₃₋₆₃₁ complex was compared with that of the full length His-NS3₁₋₆₃₁-NS4A₁₋₅₄ complex under their corresponding optimal buffer conditions. The double stranded RNA substrate (Oligos, Etc., Inc. Wilsonville, OR) used in the assay contained a template 5'-GCU CGC CCG GGG AUC CUC UAG GAA UAC ACG UUC GAU-3' (SEQ ID NO: 121) annealed to a primer 5'-CUA GAG GAU CCC CGG GGC AGC CCU AUA GUG AGU CGU-3' (complementary

15

20

25

30

35

sequences of the template and the primer are underlined). This substrate is end-labeled with ^{39}P using T4 polynucleotide kinase.

The assay conditions for the covalent His-NS4A₂₁₋₃₂-GSGS-NS3₃₋₆₃₁ complex were 100 mM MOPS [pH 7.0], 0.5 mM MgCl₂, 2 mM ATP, 0.5 mM DTT, 100 mg/ml BSA, 2% dimethylsulfoxide (DMSO) and 1 U RNase inhibitor (5 prime->3 prime, Inc., Boulder, CO). For the full length His-NS3 $_{1-631}$ /NS4A $_{1-54}$ complex, the assay conditions were 100 m M PIPES [pH 6.0], 1 mM MgCl₂, 2 mM ATP, 0.6 mM DTT, 100 mg/ml BSA and 1 U RNase inhibitor. In both reactions, 0.5 nM double stranded RNA substrate in a final volume of 50 ml was used. The reaction was carried out at 37 ∞C for 1 h and terminated by an addition of 10 ml of 0.5 M EDTA. The released primer was captured using 60 ml of 100 nM biotinylated capture oligomer (5'-biotin-GCT-CGC-CCG-GGG-ATC-CTC-TAG-3') (Gibco/BRL, Grand Island, NY) (SEQ ID NO: 123) in 2X hybridization buffer (40 mM HEPES [pH 7.3], 2M NaCl, 2 mg/ml BSA) at 37 ∞C for 1 h. The primer-oligomer complex was retained by Streptavidin coated SPA beads (SPA, Amersham Life Science Inc., Arlington Height, IL), filtered and washed thoroughly with wash buffer (20 mM HEPES [pH 7.3], 15 mM NaCl, 1.5 mM sodium citrate and 0.05% SDS). The amount of the released labeled primer was quantified using a TopCount reader (Packard A991200, Meriden, CT).

As shown in Fig. 6, the covalent His-NS4A_{21,32}-GSGS-NS3₃₋₆₃₁ displayed nucleic acid unwinding activity which was proportional to the concentration of enzyme. In the linear range of the assay for both enzymes (1 - 10 pM), about 5 - 6 fold more product was released by the His-NS4A_{21,32}-GSGS-NS3₃₋₆₃₁ than that from an equivalent concentration of full length His-NS3₁₋₆₃₁/NS4A₁₋₅₄ complex. In addition, 10 fold less covalent His-NS4A_{21,32}-GSGS-NS3₃₋₆₃₁ complex was required to yield a similar percentage of unwound products compared with the full length His-NS3₁₋₆₃₁/NS4A₁₋₅₄ complex in the corresponding reactions.

The nucleic acid unwinding activity associated with the recombinant covalent His-NS4A $_{21.32}$ -GSGS-NS3 $_{3.631}$ complex is useful for screening inhibitors of this function. For antiviral screening, compounds were tested at concentrations of less than 40 mM in the assay conditions as described above except that 0.3 nM of the double stranded RNA substrate and 20 pM of the covalent His-NS4A $_{21.32}$ -GSGS-NS3 $_{3.631}$ complex were used in a reaction which was carried out at room temperature for 30 minutes. The inhibition of the enzyme was

10

15

20

25

monitored by a decrease in the level of released labeled primer as reflected in fewer counts in the capture assay. IC_{50} of the inhibitory compounds was determined as the concentration of the compounds required to inhibit 50% of the unwinding activity.

EXAMPLE 8

Determination of ATPase activity

ATPase activity of the covalent His-NS4A_{21.32}-GSGS-NS3₃₋₆₃₁ complex (SEQ ID NO: 4) was monitored by direct measurement of [a-³²P]ATP hydrolysis using thin layer chromatography. The enzyme was incubated with 1 mM ATP mixed with [a-³²P]ATP (3000 Ci/mmol, approximately 0.5 mCi per reaction) in a reaction buffer containing 50 mM HEPES [pH 7.3], 10 mM KCl, 0.5 mM DTT, 100 mg/ml bovine serum albumin, fraction V (BSA), 1 mM MgCl₂ in the presence or absence of 1 mM polyuridylic acid (poly U) (Pharmacia, Piscataway, NJ) in a final volume of 10 ml. The reaction was carried out at 37 ∞ C for 1 h and terminated by an addition of 1 ml of 0.5 M EDTA. Half a microliter of the reaction mix was spotted onto a polyethyleneimine-cellulose sheet (SA Scientific Adsorbents Inc., Atlanta, GA) and developed by ascending chromatography in 0.375 M potassium phosphate buffer [pH 3.5]. The cellulose sheet was dried and quantified with a Storm 860 PhosphoImager (Molecular Dynamics, Sunnyvale, CA).

The covalent His-NS4A $_{2l\cdot32}$ -GSGS-NS3 $_{3\cdot611}$ complex exhibited poly U dependent ATPase activity which was proportional to the concentration of the enzyme. The ATP hydrolysis (8 - 13 fold increase) was enhanced in the presence of poly U at all enzyme concentrations examined (see Figure 7). Only minimal ATP hydrolysis was observed in the absence of poly U.

The presence of ATPase activity in this covalent ${\rm His\text{-}NS4A_{21\cdot32}^{-}}$ 30 GSGS-NS3₃₋₆₃₁ complex makes it suitable for screening inhibitors against HCV helicase.

25

30

WE CLAIM:

- A covalent HCV NS4A-NS3 complex comprising the
 central hydrophobic domain of native HCV NS4A peptide, a linker, and the HCV NS3 serine protease domain, wherein the hydrophobic domain of native HCV NS4A peptide is tethered by the linker to the amino terminus of the HCV NS3 protease domain.
- The covalent HCV NS4A-NS3 complex of claim 1, wherein the linker comprises at least about 4 amino acid residues.
 - 3. The covalent HCV NS4A-NS3 complex of claim 2, wherein the linker consists essentially of 4-6 amino acid residues.

 $4. \qquad \text{The covalent HCV NS4A-NS3 complex of claim 3, wherein the linker consists essentially of about 4 amino acid residues.}$

- The covalent HCV NS4A-NS3 complex of claim 4, wherein
 the linker has a sequence defined by SEQ ID NO: 21 or SEQ ID NO: 22.
 - The covalent HCV NS4A-NS3 complex of claim 5, having an amino acid sequence selected from the group consisting of SEQ ID NOs: 1-20.
 - 7. The covalent HCV NS4A-NS3 complex of claim 1 which is modified by replacement of one or more hydrophobic amino acid residues at position 17 or 18 of the HCV NS3 serine protease domain with a hydrophilic amino acid residue.
 - 8. The covalent HCV NS4A-NS3 complex of claim 7 in which one or more isoleucine residues at position 17 or 18 of the HCV NS3 serine protease domain is replaced by a lysine residue.
- 9. The covalent HCV NS4A-NS3 complex of claim 8, having an amino acid sequence selected from the group consisting of SEQ ID NOs: 2-4, 6-8, 10, 12-14 and 16-18.

10

15

20

- 10. The covalent HCV NS4A-NS3 complex of claim 1 which is modified by replacement of a serine residue at position 139 of the HCV NS3 serine protease domain with an alanine residue.
- 11. The covalent HCV NS4A-NS3 complex of claim 10, having an amino acid sequence selected from the group consisting of SEQ ID NOs: 5-8, 15-18 and 20.
- 12. A nucleic acid encoding a covalent HCV NS4A-NS3 complex, which covalent HCV NS4A-NS3 complex comprises the central hydrophobic domain of native HCV NS4A peptide, a linker, and the HCV NS3 serine protease domain, wherein the hydrophobic domain of native HCV NS4A peptide is tethered by the amino acid linker to the amino terminus of the HCV NS3 protease domain.
- 13. The nucleic acid of claim 12, wherein the linker comprises a least about 4 amino acid residues.
- 14. The nucleic acid of claim 13, wherein the linker consists essentially of 4-6 amino acid residues.
 - The nucleic acid of claim 14, wherein the linker consists essentially of 4 amino acid residues.
- 16. $\;$ The nucleic acid of claim 15, wherein the amino acid linker has a sequence defined by SEQ ID NO: 21 or SEQ ID NO: 22.
- The nucleic acid of claim 16, which encodes a covalent
 HCV NS4A-NS3 complex having an amino acid sequence selected from the group consisting of SEQ ID NOs: 1-20.
- The nucleic acid of claim 12, which encodes a covalent HCV NS4A-NS3 complex which is modified by replacement of one or more
 hydrophobic amino acid residues at position 17 or 18 of the HCV NS3 serine protease domain with a hydrophilic amino acid residue.

- 19. The nucleic acid of claim 18 which encodes a covalent HCV NS4A-NS3 complex in which one or more isoleucine residues at position 17 or 18 of the HCV NS3 serine protease domain are replaced by a lysine residue.
- 20. The nucleic acid of claim 19, which encodes a covalent HCV NS4A-NS3 complex having an amino acid sequence selected from the group consisting of SEQ ID NOs: 2-4, 6-8, 10, 12-14 and 16-18.
- 10 21. The nucleic acid of claim 12, which encodes a covalent HCV NS4A-NS3 complex which is modified by replacement of a serine residue at position 139 of the HCV NS3 serine protease domain with an alanine residue.
- 15 22. The nucleic acid of claim 21, which encodes a covalent HCV NS4A-NS3 complex having an amino acid sequence selected from the group consisting of SEQ ID NOs: 5-8, 15-18 and 20.
- A recombinant vector comprising the nucleic acid of claim
 12, which vector is capable of directing expression of the nucleic acid.
 - 24. A host cell comprising the recombinant vector of claim 23.
- 25. A method for making a covalent HCV NS4A-NS3 complex 25 comprising culturing the host cell of claim 24 under conditions in which the nucleic acid or vector is expressed.
 - 26. A method for identifying an HCV NS3 protease inhibitor, comprising (a) contacting a covalent HCV NS4A-NS3 complex of claim 1 with a peptide substrate and a suspected protease inhibitor under conditions in which proteolysis can occur; and (b) detecting whether the covalent HCV NS4-NS3 complex has cleaved the substrate.
- 27. A method for identifying an inhibitor of the nucleic acid unwinding activity of an HCV NS3 helicase, comprising (a) contacting a covalent HCV NS4A-NS3 complex of SEQ ID NO: 4, 12-19 or 20 with a double stranded RNA substrate and a suspected inhibitor under conditions in which unwinding of the substrate can occur; and (b)

detecting whether and the extent to which the covalent HCV NS4-NS3 complex has unwound the substrate.

28. A method for identifying an inhibitor of an HCV NS3 helicase, comprising (a) contacting a covalent HCV NS4A-NS3 complex of SEQ ID NO: 4, 12-19 or 20 with ATP and a suspected inhibitor under conditions in which ATP hydrolysis can occur; and (b) detecting whether the covalent HCV NS4-NS3 complex has exhibited ATPase activity.

10

ABSTRACT OF THE DISCLOSURE

Covalent HCV NS4A-NS3 complexes comprising the

5 central hydrophobic domain of native HCV NS4A peptide, a linker, and
the HCV NS3 serine protease domain, wherein the hydrophobic domain
of native HCV NS4A peptide is tethered by the linker to the amino
terminus of the HCV NS3 protease domain.

SEQUENCE LISTING

- (1) GENERAL INFORMATION:
 - (i) APPLICANT: Malcolm, Bruce Taremi, Shahriar S. Weber, Patricia Yao, Nanhua
 - (ii) TITLE OF INVENTION: Covalent Complexes of Hepatitis C Virus NS3 Protease and NS4A Cofactor Peptide
 - (iii) NUMBER OF SEQUENCES: 123
 - (iv) CORRESPONDENCE ADDRESS:
 - (A) ADDRESSEE: Schering-Plough Corp.
 - (B) STREET: 2000 Galloping Hill Road
 - (C) CITY: Kenilworth
 - (D) STATE: New Jersey
 - (E) COUNTRY: USA
 - (F) ZIP: 07030
 - (v) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Floppy disk
 - (B) COMPUTER: Power Macintosh
 - (C) OPERATING SYSTEM: 8.0.1
 (D) SOFTWARE: Microsoft Word 6.0.1
 - (vi) CURRENT APPLICATION DATA:
 - (A) APPLICATION NUMBER: US
 - (B) FILING DATE:
 - (C) CLASSIFICATION:
 - (viii) ATTORNEY/AGENT INFORMATION:
 - (A) NAME: McLaughlin, Jaye P.
 - (B) REGISTRATION NUMBER: 41,211
 - (C) REFERENCE/DOCKET NUMBER: JB0800P2
 - (ix) TELECOMMUNICATION INFORMATION:
 - (A) TELEPHONE: (908)298-5056
 - (B) TELEFAX: (908)298-5388
- (2) INFORMATION FOR SEO ID NO:1:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 217 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: protein
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

Met Gly Ser Ser His His His His His Ser Ser Gly Leu Val Pro

1 5 10 15

Arg Gly Ser His Met Gly Ser Val Val Ile Val Gly Arg Ile Ile Leu $20 \\ 25 \\ 30$

Ser Gly Ser Gly Ser Ile Thr Ala Tyr Ser Gln Gln Thr Arg Gly Leu \$35\$

Leu Gly Cys Ile Ile Thr Ser Leu Thr Gly Arg Asp Lys Asn Gln Val $_{50}$ $_{60}$

Glu Gly Glu Val Gln Val Val Ser Thr Ala Thr Gln Ser Phe Leu Ala 65 707075

Lys Thr Leu Ala Gly Pro Lys Gly Pro Ile Thr Gln Met Tyr Thr Asn 100 105 110 .

Val Asp Gln Asp Leu Val Gly Trp Gln Ala Pro Pro Gly Ala Arg Ser 115 120 125

Leu Thr Pro Cys Thr Cys Gly Ser Ser Asp Leu Tyr Leu Val Thr Arg $130 \ \ 135 \ \ 140$

His Ala Asp Val Ile Pro Val Arg Arg Gly Asp Ser Arg Gly Ser 145 150 155 160

Leu Leu Ser Pro Arg Pro Val Ser Tyr Leu Lys Gly Ser Ser Gly Gly 165 170 175

Pro Leu Cus Pro Ser Gly His Ala Val Gly Ile Phe Arg Ala Ala 180 185 190

Val Cys Thr Arg Gly Val Ala Lys Ala Val Asp Phe Val Pro Val Glu 195 200 205

Ser Met Glu Thr Thr Met Arg Ser 210

(2) INFORMATION FOR SEO ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 217 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Gly Ser Ser His His His His His Ser Ser Gly Leu Val Pro $\frac{1}{5}$

Arg Gly Ser His Met Gly Ser Val Val Ile Val Gly Arg Ile Ile Leu

Ser	Gly	Ser	Gly	Ser	Ile	Thr	Ala	Tyr	Ser	${\tt Gln}$	Gln	Thr	Arg	Gly	Leu
		35					40					45			

Leu Gly Cys Lys Ile Thr Ser Leu Thr Gly Arg Asp Lys Asn Gln Val

Glu Gly Glu Val Gln Val Val Ser Thr Ala Thr Gln Ser Phe Leu Ala 70

Thr Cys Val Asn Gly Val Cys Trp Thr Val Tyr His Gly Ala Gly Ser 90

Lys Thr Leu Ala Gly Pro Lys Gly Pro Ile Thr Gln Met Tyr Thr Asn 105

Val Asp Gln Asp Leu Val Gly Trp Gln Ala Pro Pro Gly Ala Arg Ser 120

Leu Thr Pro Cys Thr Cys Gly Ser Ser Asp Leu Tyr Leu Val Thr Arg 130

His Ala Asp Val Ile Pro Val Arg Arg Gly Asp Ser Arg Gly Ser 150 155

Leu Leu Ser Pro Arg Pro Val Ser Tyr Leu Lys Gly Ser Ser Gly Gly

Pro Leu Cys Pro Ser Gly His Ala Val Gly Ile Phe Arg Ala Ala

Val Cys Thr Arg Gly Val Ala Lys Ala Val Asp Phe Val Pro Val Glu 200

Ser Met Glu Thr Thr Met Arg Ser 210 215

i de 03

11

10.

1 100

133 :500 D

(13

(2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 217 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Met Gly Ser Ser His His His His His Ser Ser Gly Leu Val Pro

Arg Gly Ser His Met Gly Ser Val Val Ile Val Gly Arg Ile Ile Leu 20

Ser Gly Ser Gly Ser Ile Thr Ala Tyr Ser Gln Gln Thr Arg Gly Leu

35 40 45

Leu Gly Cys Ile Lys Thr Ser Leu Thr Gly Arg Asp Lys Asn Gln Val

Thr Cys Val Asn Gly Val Cys Trp Thr Val Tyr His Gly Ala Gly Ser $85 \hspace{1.5cm} 90 \hspace{1.5cm} 95$

Lys Thr Leu Ala Gly Pro Lys Gly Pro Ile Thr Gln Met Tyr Thr Asn $100 \\ 105 \\ 110$

Val Asp Gln Asp Leu Val Gly Trp Gln Ala Pro Pro Gly Ala Arg Ser 115 120 125

Leu Thr Pro Cys Thr Cys Gly Ser Ser Asp Leu Tyr Leu Val Thr Arg 130 135 140

His Ala Asp Val Ile Pro Val Arg Arg Arg Gly Asp Ser Arg Gly Ser 145 \$150\$

Leu Leu Ser Pro Arg Pro Val Ser Tyr Leu Lys Gly Ser Ser Gly Gly 175 \$175

Pro Leu Cys Pro Ser Gly His Ala Val Gly Ile Phe Arg Ala Ala 180 \$180\$

Val Cys Thr Arg Gly Val Ala Lys Ala Val Asp Phe Val Pro Val Glu 195 200 205

Ser Met Glu Thr Thr Met Arg Ser

(2) INFORMATION FOR SEQ ID NO:4:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 217 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Met Gly Ser Ser His His His His His Ser Ser Gly Leu Val Pro 1

Arg Gly Ser His Met Gly Ser Val Val Ile Val Gly Arg Ile Ile Leu
20 25 30

Ser Gly Ser Gly Ser Ile Thr Ala Tyr Ser Gln Gln Thr Arg Gly Leu $35 \ \ \, 40 \ \ \, 45$

Leu Gly Cys Lys Lys Thr Ser Leu Thr Gly Arg Asp Lys Asn Gln Val

50 55 60

Glu Gly Glu Val Gln Val Val Ser Thr Ala Thr Gln Ser Phe Leu Ala 65 70 75 80

Thr Cys Val Asn Gly Val Cys Trp Thr Val Tyr His Gly Ala Gly Ser 85 90 95

Lys Thr Leu Ala Gly Pro Lys Gly Pro Ile Thr Gln Met Tyr Thr Asn $100 \hspace{1.5cm} 105 \hspace{1.5cm} 110 \hspace{1.5cm}$

Val Asp Gln Asp Leu Val Gly Trp Gln Ala Pro Pro Gly Ala Arg Ser 115 120 125

Leu Thr Pro Cys Thr Cys Gly Ser Ser Asp Leu Tyr Leu Val Thr Arg 130 135 140

His Ala Asp Val Ile Pro Val Arg Arg Arg Gly Asp Ser Arg Gly Ser 145 \$150\$

Leu Leu Ser Pro Arg Pro Val Ser Tyr Leu Lys Gly Ser Ser Gly Gly 165 170 175

Pro Leu Leu Cys Pro Ser Gly His Ala Val Gly Ile Phe Arg Ala Ala 180 185 190

Val Cys Thr Arg Gly Val Ala Lys Ala Val Asp Phe Val Pro Val Glu 195 200 205

Ser Met Glu Thr Thr Met Arg Ser * 210 215

(2) INFORMATION FOR SEO ID NO:5:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 216 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Met Gly Ser Ser His His His His His Ser Ser Gly Leu Val Pro

1 5 10 15

Arg Gly Ser His Met Gly Ser Val Val Ile Val Gly Arg Ile Ile Leu $20 \hspace{1.5cm} 25 \hspace{1.5cm} 30 \hspace{1.5cm}$

Ser Gly Ser Gly Ser Ile Thr Ala Tyr Ser Gln Gln Thr Arg Gly Leu $35 \hspace{1cm} 40 \hspace{1cm} 45 \hspace{1cm}$

Leu Gly Cys Ile Ile Thr Ser Leu Thr Gly Arg Asp Lys Asn Gln Val 50 60

Glu Gly Glu Val Gln Val Val Ser Thr Ala Thr Gln Ser Phe Leu Ala

Thr Cys Val Asn Gly Val Cys Trp Thr Val Tyr His Gly Ala Gly Ser $85 \hspace{0.25in} 90 \hspace{0.25in} 95$

Val Asp Gln Asp Leu Val Gly Trp Gln Ala Pro Pro Gly Ala Arg Ser 115 120 125

Leu Thr Pro Cys Thr Cys Gly Ser Ser Asp Leu Tyr Leu Val Thr Arg 130 135 140

His Ala Asp Val Ile Pro Val Arg Arg Arg Gly Asp Ser Arg Gly Ser 145 \$150\$

Leu Leu Ser Pro Arg Pro Val Ser Tyr Leu Lys Gly Ser Ala Gly Gly 165 170 175 .

Pro Leu Cys Pro Ser Gly His Ala Val Gly Ile Phe Arg Ala Ala 180 185 190

Val Cys Thr Arg Gly Val Ala Lys Ala Val Asp Phe Val Pro Val Glu 195 200 205

Ser Met Glu Thr Thr Met Arg Ser 210 215

(2) INFORMATION FOR SEQ ID NO:6:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 216 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Met Gly Ser Ser His His His His His Ser Ser Gly Leu Val Pro $1 \hspace{1cm} 5 \hspace{1cm} 10 \hspace{1cm} 15$

Arg Gly Ser His Met Gly Ser Val Val Ile Val Gly Arg Ile Ile Leu $20 \\ 25 \\ 30$

Ser Gly Ser Gly Ser Ile Thr Ala Tyr Ser Gln Gln Thr Arg Gly Leu $35 \hspace{1cm} 40 \hspace{1cm} 45$

Leu Gly Cys Lys Ile Thr Ser Leu Thr Gly Arg Asp Lys Asn Gln Val50

Glu Gly Glu Val Gln Val Val Ser Thr Ala Thr Gln Ser Phe Leu Ala 65 70 75 80

Thr Cys Val Asn Gly Val Cys Trp Thr Val Tyr His Gly Ala Gly Ser

1

F

m

85 90 95

Lys Thr Leu Ala Gly Pro Lys Gly Pro Ile Thr Gln Met Tyr Thr Asn 100 105 110

Val Asp Gln Asp Leu Val Gly Trp Gln Ala Pro Pro Gly Ala Arg Ser 115 120 125

Leu Thr Pro Cys Thr Cys Gly Ser Ser Asp Leu Tyr Leu Val Thr Arg 130 135

His Ala Asp Val Ile Pro Val Arg Arg Arg Gly Asp Ser Arg Gly Ser 145 \$150\$

Leu Leu Ser Pro Arg Pro Val Ser Tyr Leu Lys Gly Ser Ala Gly Gly 165 170 175

Pro Leu Cys Pro Ser Gly His Ala Val Gly Ile Phe Arg Ala Ala 180 185 190 .

Val Cys Thr Arg Gly Val Ala Lys Ala Val Asp Phe Val Pro Val Glu
195 200 205

Ser Met Glu Thr Thr Met Arg Ser 210 215

- (2) INFORMATION FOR SEQ ID NO:7:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 217 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: protein
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Met Gly Ser Ser His His His His His Ser Ser Gly Leu Val Pro

1 5 10 15

Arg Gly Ser His Met Gly Ser Val Val Ile Val Gly Arg Ile Ile Leu $20 \\ 25 \\ 30$

Ser Gly Ser Gly Ser Ile Thr Ala Tyr Ser Gln Gln Thr Arg Gly Leu 35 40 45

Leu Gly Cys Ile Lys Thr Ser Leu Thr Gly Arg Asp Lys Asn Gln Val50 60

Glu Gly Glu Val Gln Val Val Ser Thr Ala Thr Gln Ser Phe Leu Ala 65 70 75 80

Thr Cys Val Asn Gly Val Cys Trp Thr Val Tyr His Gly Ala Gly Ser 85 90 95

Lys Thr Leu Ala Gly Pro Lys Gly Pro Ile Thr Gln Met Tyr Thr Asn

100 105 110

Val Asp Gln Asp Leu Val Gly Trp Gln Ala Pro Pro Gly Ala Arg Ser 115 120 125

Leu Thr Pro Cys Thr Cys Gly Ser Ser Asp Leu Tyr Leu Val Thr Arg 130 135 140

His Ala Asp Val Ile Pro Val Arg Arg Arg Gly Asp Ser Arg Gly Ser 145 \$150\$

Leu Leu Ser Pro Arg Pro Val Ser Tyr Leu Lys Gly Ser Ala Gly Gly
165 170 175

Pro Leu Cys Pro Ser Gly His Ala Val Gly Ile Phe Arg Ala Ala 180 185 190

Val Cys Thr Arg Gly Val Ala Lys Ala Val Asp Phe Val Pro Val Glu
195 200 205 .

Ser Met Glu Thr Thr Met Arg Ser * 210 215

(2) INFORMATION FOR SEQ ID NO:8:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 217 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Met Gly Ser Ser His His His His His Ser Ser Gly Leu Val Pro $1 \hspace{1cm} 5 \hspace{1cm} 10 \hspace{1cm} 15$

Arg Gly Ser His Met Gly Ser Val Val Ile Val Gly Arg Ile Ile Leu 20 25 30

Ser Gly Ser Gly Ser Ile Thr Ala Tyr Ser Gln Gln Thr Arg Gly Leu $35 \hspace{1cm} 40 \hspace{1cm} 45 \hspace{1cm}$

Leu Gly Cys Lys Lys Thr Ser Leu Thr Gly Arg Asp Lys Asn Gln Val 50 60

Glu Gly Glu Val Gln Val Val Ser Thr Ala Thr Gln Ser Phe Leu Ala 65 70 75 80

Thr Cys Val Asn Gly Val Cys Trp Thr Val Tyr His Gly Ala Gly Ser

Lys Thr Leu Ala Gly Pro Lys Gly Pro Ile Thr Gln Met Tyr Thr Asn $100 \hspace{1.5cm} 105 \hspace{1.5cm} 110 \hspace{1.5cm}$

Val Asp Gln Asp Leu Val Gly Trp Gln Ala Pro Pro Gly Ala Arg Ser

Leu Thr Pro Cys Thr Cys Gly Ser Ser Asp Leu Tyr Leu Val Thr Arg 130 135 140

His Ala Asp Val Ile Pro Val Arg Arg Gly Asp Ser Arg Gly Ser 145 150 155 160

Leu Leu Ser Pro Arg Pro Val Ser Tyr Leu Lys Gly Ser Ala Gly Gly
165 170 175

Pro Leu Cys Pro Ser Gly His Ala Val Gly Ile Phe Arg Ala Ala 180 185 190

Val Cys Thr Arg Gly Val Ala Lys Ala Val Asp Phe Val Pro Val Glu
195 200 205

Ser Met Glu Thr Thr Met Arg Ser * 210 215

- (2) INFORMATION FOR SEO ID NO:9:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 217 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: protein
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

Met Gly Ser Ser His His His His His Ser Ser Gly Leu Val Pro $1 \hspace{1cm} 5 \hspace{1cm} 10 \hspace{1cm} 15$

Arg Gly Ser His Met Gly Ser Val Val Ile Val Gly Arg Ile Ile Leu 20 25 30

Ser Pro Ala Gly Gly Ile Thr Ala Tyr Ser Gln Gln Thr Arg Gly Leu 35 40 45

Leu Gly Cys Ile Ile Thr Ser Leu Thr Gly Arg Asp Lys Asn Gln Val 50 60

Glu Gly Glu Val Gln Val Val Ser Thr Ala Thr Gln Ser Phe Leu Ala 65 70 75 80

Thr Cys Val Asn Gly Val Cys Trp Thr Val Tyr His Gly Ala Gly Ser \$85\$ 90 95

Val Asp Gln Asp Leu Val Gly Trp Gln Ala Pro Pro Gly Ala Arg Ser 115 120 125 Leu Thr Pro Cys Thr Cys Gly Ser Ser Asp Leu Tyr Leu Val Thr Arg 130 135 140

His Ala Asp Val Ile Pro Val Arg Arg Gly Asp Ser Arg Gly Ser 145 150 155 160

Leu Leu Ser Pro Arg Pro Val Ser Tyr Leu Lys Gly Ser Ser Gly Gly
165 170 175

Pro Leu Leu Cys Pro Ser Gly His Ala Val Gly Ile Phe Arg Ala Ala 180 185 190

Val Cys Thr Arg Gly Val Ala Lys Ala Val Asp Phe Val Pro Val Glu . 195 200 205

Ser Met Glu Thr Thr Met Arg Ser * 210 215

- (2) INFORMATION FOR SEQ ID NO:10:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 217 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: protein
 - (xi) SEQUENCE DESCRIPTION: SEO ID NO:10:

Met Gly Ser Ser His His His His His Ser Ser Gly Leu Val Pro $1 \hspace{1cm} 5 \hspace{1cm} 10 \hspace{1cm} 15$

Arg Gly Ser His Met Gly Ser Val Val Ile Val Gly Arg Ile Ile Leu 20 25 30

Ser Pro Ala Gly Gly Ile Thr Ala Tyr Ser Gln Gln Thr Arg Gly Leu $35 \hspace{1.5cm} 40 \hspace{1.5cm} 45$

Glu Gly Glu Val Gln Val Val Ser Thr Ala Thr Gln Ser Phe Leu Ala $65 \hspace{1.5cm} 70 \hspace{1.5cm} 75 \hspace{1.5cm} 80$

Thr Cys Val Asn Gly Val Cys Trp Thr Val Tyr His Gly Ala Gly Ser 85 90 95

Lys Thr Leu Ala Gly Pro Lys Gly Pro Ile Thr Gln Met Tyr Thr Asn $100 \hspace{1.5cm} 105 \hspace{1.5cm} 110 \hspace{1.5cm}$

Val Asp Gln Asp Leu Val Gly Trp Gln Ala Pro Pro Gly Ala Arg Ser 115 120 125

Leu Thr Pro Cys Thr Cys Gly Ser Ser Asp Leu Tyr Leu Val Thr Arg 130 135 140

His Ala Asp Val Ile Pro Val Arg Arg Arg Gly Asp Ser Arg Gly Ser 145 \$150\$

Leu Leu Ser Pro Arg Pro Val Ser Tyr Leu Lys Gly Ser Ser Gly Gly $165 \hspace{1.5cm} 170 \hspace{1.5cm} 175$

Pro Leu Cys Pro Ser Gly His Ala Val Gly Ile Phe Arg Ala Ala 180 185 190

Val Cys Thr Arg Gly Val Ala Lys Ala Val Asp Phe Val Pro Val Glu 195 200 205

Ser Met Glu Thr Thr Met Arg Ser * 210 215

- (2) INFORMATION FOR SEQ ID NO:11:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 665 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: protein
 - (xi) SEQUENCE DESCRIPTION: SEO ID NO:11:

Met Gly Ser Ser His His His His His Ser Ser Gly Leu Val Pro $1 \hspace{1cm} 5 \hspace{1cm} 10 \hspace{1cm} 15$

Arg Gly Ser His Met Gly Ser Val Val Ile Val Gly Arg Ile Ile Leu $20 \\ 25 \\ 30$

Ser Gly Ser Gly Ser Ile Thr Ala Tyr Ser Gln Gln Thr Arg Gly Leu $35 \hspace{1cm} 40 \hspace{1cm} 45$

Glu Gly Glu Val Gln Val Val Ser Thr Ala Thr Gln Ser Phe Leu Ala 65 70 75 80

Thr Cys Val Asn Gly Val Cys Trp Thr Val Tyr His Gly Ala Gly Ser 85 90 95

Lys Thr Leu Ala Gly Pro Lys Gly Pro Ile Thr Gln Met Tyr Thr Asn $100 \hspace{1.5cm} 105 \hspace{1.5cm} 110 \hspace{1.5cm}$

Val Asp Gln Asp Leu Val Gly Trp Gln Ala Pro Pro Gly Ala Arg Ser 115 120 125

Leu Thr Pro Cys Thr Cys Gly Ser Ser Asp Leu Tyr Leu Val Thr Arg 130 135 140

His Ala Asp Val Ile Pro Val Arg Arg Arg Gly Asp Ser Arg Gly Ser 145 150 155 160

- Leu Leu Ser Pro Arg Pro Val Ser Tyr Leu Lys Gly Ser Ser Gly Gly 165 170 175
- Pro Leu Cys Pro Ser Gly His Ala Val Gly Ile Phe Arg Ala Ala 180 185 190
- Val Cys Thr Arg Gly Val Ala Lys Ala Val Asp Phe Val Pro Val Glu 195 200 205
- Ser Met Glu Thr Thr Met Arg Ser Pro Val Phe Thr Asp Asn Ser Ser 210 215 220
- Pro Pro Ala Val Pro Gln Ser Phe Gln Val Ala His Leu His Ala Pro 225 230 230
- Thr Gly Ser Gly Lys Ser Thr Lys Val Pro Ala Ala Tyr Ala Ala Gln 245 250
- Gly Tyr Lys Val Leu Val Leu Asn Pro Ser Val Ala Ala Thr Leu Gly $260 \hspace{1cm} 265 \hspace{1cm} 265 \hspace{1cm} 270 \hspace{1cm}$
- Phe Gly Ala Tyr Met Ser Lys Ala His Gly Ile Asp Pro Asn Ile Arg 275 280 285
- Thr Gly Val Arg Thr Ile Thr Thr Gly Ala Pro Val Thr Tyr Ser Thr 290 295 300
- Tyr Gly Lys Phe Leu Ala Asp Gly Gly Cys Ser Gly Gly Ala Tyr Asp 305 310 315 320
- Ile Ile Ile Cys Asp Glu Cys His Ser Thr Asp Ser Thr Thr Ile Leu 325 330 335
- Gly Ile Gly Thr Val Leu Asp Gln Ala Glu Thr Ala Gly Ala Arg Leu $340 \hspace{1.5cm} 345 \hspace{1.5cm} 350$
- Val Val Leu Ala Thr Ala Thr Pro Pro Gly Ser Val Thr Val Pro His 355 360 365
- Pro Asn Ile Glu Glu Val Ala Leu Ser Asn Thr Gly Glu Ile Pro Phe $370 \ \ 375 \ \ 380$
- Tyr Gly Lys Ala Ile Pro Ile Glu Ala Ile Arg Gly Gly Arg His Leu 385 390 395 400
- Ile Phe Cys His Ser Lys Lys Lys Cys Asp Glu Leu Ala Ala Lys Leu 405 410 415
- Ser Gly Leu Gly Ile Asn Ala Val Ala Tyr Tyr Arg Gly Leu Asp Val \$420\$ \$425\$
- Ser Val Ile Pro Thr Ile Gly Asp Val Val Val Val Ala Thr Asp Ala 435 440 445
- Leu Met Thr Gly Tyr Thr Gly Asp Phe Asp Ser Val Ile Asp Cys Asn 450 455
- Thr Cys Val Thr Gln Thr Val Asp Phe Ser Leu Asp Pro Thr Phe Thr

465					470					475					480
Ile	Glu	Thr	Thr	Thr 485	Val	Pro	Gln	Asp	Ala 490	Val	Ser	Arg	Ser	Gln 495	Arg
Arg	Gly	Arg	Thr 500	Gly	Arg	Gly	Arg	Arg 505	Gly	Ile	Tyr	Arg	Phe 510	Val	Thr
Pro	Gly	Glu 515	Arg	Pro	Ser	Gly	Met 520	Phe	Asp	Ser	Ser	Val 525	Leu	Cys	Glu
Cys	Tyr 530	Asp	Ala	Gly	Cys	Ala 535	Trp	Tyr	Glu	Leu	Thr 540	Pro	Ala	Glu	Thr
Ser 545	Val	Arg	Leu	Arg	Ala 550	Tyr	Leu	Asn	Thr	Pro 555	Gly	Leu	Pro	Val	Суs 560
Gln	Asp	His	Leu	Glu 565	Phe	Trp	Glu	Ser	Val 570	Phe	Thr	Gly	Leu	Thr 575	His
Ile	Asp	Ala	His 580	Phe	Leu	Ser	Gln	Thr 585	Lys	Gln	Ala	Gly	Asp 590	Asn	Phe
Pro	Tyr	Leu 595	Val	Ala	Tyr	Gln	Ala 600	Thr	Val	Cys	Ala	Arg 605	Ala	Gln	Ala
Pro	Pro 610	Pro	Ser	Trp	Asp	Gln 615	Met	Trp	Lys	Cys	Leu 620	Ile	Arg	Leu	Lys
Pro 625	Thr	Leu	His	Gly	Pro 630	Thr	Pro	Leu	Leu	Tyr 635	Arg	Leu	Gly	Ala	Val 640
Gln	Asn	Glu	Val	Thr 645	Leu	Thr	His	Pro	Ile 650	Thr	Lys	Tyr	Ile	Met 655	Ala

(2) INFORMATION FOR SEQ ID NO:12:

Cys Met Ser Ala Asp Leu Glu Val Val 660 665

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 665 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

Met Gly Ser Ser His His His His His Ser Ser Gly Leu Val Pro

1 10 15

Arg Gly Ser His Met Gly Ser Val Val Ile Val Gly Arg Ile Ile Leu $20 \\ 25 \\ 30$

Ser Gly Ser Gly Ser Ile Thr Ala Tyr Ser Gln Gln Thr Arg Gly Leu

10

Leu	Gly 50	Cys	Lys	Ile	Thr	Ser 55	Leu	Thr	Gly	Arg	Asp 60	Lys	Asn	Gln	Val

40

35

Glu Gly Glu Val Gln Val Val Ser Thr Ala Thr Gln Ser Phe Leu Ala $65 \\ 70 \\ 75 \\ 80$

Thr Cys Val Asn Gly Val Cys Trp Thr Val Tyr His Gly Ala Gly Ser $85 \hspace{1cm} 90 \hspace{1cm} 95$

Lys Thr Leu Ala Gly Pro Lys Gly Pro Ile Thr Gln Met Tyr Thr Asn $100 \hspace{1.5cm} 105 \hspace{1.5cm} 110 \hspace{1.5cm}$

Val Asp Gln Asp Leu Val Gly Trp Gln Ala Pro Pro Gly Ala Arg Ser 115 120 125

Leu Thr Pro Cys Thr Cys Gly Ser Ser Asp Leu Tyr Leu Val Thr Arg 130 135 140 .

His Ala Asp Val Ile Pro Val Arg Arg Arg Gly Asp Ser Arg Gly Ser 145 150 155 160

Leu Leu Ser Pro Arg Pro Val Ser Tyr Leu Lys Gly Ser Ser Gly Gly 165 170 175

Pro Leu Cus Pro Ser Gly His Ala Val Gly Ile Phe Arg Ala Ala 180 \$185\$

Val Cys Thr Arg Gly Val Ala Lys Ala Val Asp Phe Val Pro Val Glu 195 200 205

Ser Met Glu Thr Thr Met Arg Ser Pro Val Phe Thr Asp Asn Ser Ser 210 \$215\$

Pro Pro Ala Val Pro Gln Ser Phe Gln Val Ala His Leu His Ala Pro 225 230 235 240

Thr Gly Ser Gly Lys Ser Thr Lys Val Pro Ala Ala Tyr Ala Ala Gln 245 255

Gly Tyr Lys Val Leu Val Leu Asn Pro Ser Val Ala Ala Thr Leu Gly $260 \hspace{1.5cm} 265 \hspace{1.5cm} 270 \hspace{1.5cm}$

Phe Gly Ala Tyr Met Ser Lys Ala His Gly Ile Asp Pro Asn Ile Arg 275 280 285

Thr Gly Val Arg Thr Ile Thr Thr Gly Ala Pro Val Thr Tyr Ser Thr 290 295 300

Tyr Gly Lys Phe Leu Ala Asp Gly Gly Cys Ser Gly Gly Ala Tyr Asp 305 310 315 320

Ile Ile Ile Cys Asp Glu Cys His Ser Thr Asp Ser Thr Thr Ile Leu 325 330 335

Gly Ile Gly Thr Val Leu Asp Gln Ala Glu Thr Ala Gly Ala Arg Leu $340 \hspace{1.5cm} 345 \hspace{1.5cm} 350$

Val Val Leu Ala Thr Ala Thr Pro Pro Gly Ser Val Thr Val Pro His 355 360 365

Pro Asn Ile Glu Glu Val Ala Leu Ser Asn Thr Gly Glu Ile Pro Phe $370 \ \ 375 \ \ 380$

Tyr Gly Lys Ala Ile Pro Ile Glu Ala Ile Arg Gly Gly Arg His Leu 385 390 395 400

Ile Phe Cys His Ser Lys Lys Lys Cys Asp Glu Leu Ala Ala Lys Leu 405 410 415

Ser Gly Leu Gly Ile Asn Ala Val Ala Tyr Tyr Arg Gly Leu Asp Val \$420\$

Ser Val Ile Pro Thr Ile Gly Asp Val Val Val Val Ala Thr Asp Ala 435 440 445

Leu Met Thr Gly Tyr Thr Gly Asp Phe Asp Ser Val Ile Asp Cys Asn $450 \ \ 455 \ \ 460 \ \ \$

Thr Cys Val Thr Gln Thr Val Asp Phe Ser Leu Asp Pro Thr Phe Thr 465 . 470 475 480

Arg Gly Arg Thr Gly Arg Gly Arg Gly Ile Tyr Arg Phe Val Thr $500 \hspace{1.5cm} 505 \hspace{1.5cm} 510 \hspace{1.5cm}$

Pro Gly Glu Arg Pro Ser Gly Met Phe Asp Ser Ser Val Leu Cys Glu 515 520 525

Cys Tyr Asp Ala Gly Cys Ala Trp Tyr Glu Leu Thr Pro Ala Glu Thr 530 535 540

Ser Val Arg Leu Arg Ala Tyr Leu Asn Thr Pro Gly Leu Pro Val Cys 545 550 550

Gln Asp His Leu Glu Phe Trp Glu Ser Val Phe Thr Gly Leu Thr His 565 570 575

Ile Asp Ala His Phe Leu Ser Gln Thr Lys Gln Ala Gly Asp Asn Phe $580 \hspace{1.5cm} 585 \hspace{1.5cm} 590 \hspace{1.5cm}$

Pro Tyr Leu Val Ala Tyr Gln Ala Thr Val Cys Ala Arg Ala Gln Ala 595 600 605

Pro Pro Pro Ser Trp Asp Gln Met Trp Lys Cys Leu Ile Arg Leu Lys 610 615 620

Pro Thr Leu His Gly Pro Thr Pro Leu Leu Tyr Arg Leu Gly Ala Val 625 630 -635 640

. Gln Asn Glu Val Thr Leu Thr His Pro Ile Thr Lys Tyr Ile Met Ala $645 \hspace{1.5cm} 650 \hspace{1.5cm} 655$

Cys Met Ser Ala Asp Leu Glu Val Val
660

- (2) INFORMATION FOR SEQ ID NO:13:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 665 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: protein
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

Met Gly Ser Ser His His His His His Ser Ser Gly Leu Val Pro

1 5 10 15

Arg Gly Ser His Met Gly Ser Val Val Ile Val Gly Arg Ile Ile Leu

Ser Gly Ser Gly Ser Ile Thr Ala Tyr Ser Gln Gln Thr Arg Gly Leu

Leu Gly Cys Ile Lys Thr Ser Leu Thr Gly Arg Asp Lys Asn Gln Val 50

Thr Cys Val Asn Gly Val Cys Trp Thr Val Tyr His Gly Ala Gly Ser $85 \hspace{1cm} 90 \hspace{1cm} 95$

Lys Thr Leu Ala Gly Pro Lys Gly Pro Ile Thr Gln Met Tyr Thr Asn $100 \\ 105 \\ 110$

Leu Thr Pro Cys Thr Cys Gly Ser Ser Asp Leu Tyr Leu Val Thr Arg 130 135 140

His Ala Asp Val Ile Pro Val Arg Arg Arg Gly Asp Ser Arg Gly Ser 145 150 155 160

Leu Leu Ser Pro Arg Pro Val Ser Tyr Leu Lys Gly Ser Ser Gly Gly 165 170 175

Pro Leu Cys Pro Ser Gly His Ala Val Gly Ile Phe Arg Ala Ala 180 185 190

Val Cys Thr Arg Gly Val Ala Lys Ala Val Asp Phe Val Pro Val Glu 195 200 205

Ser Met Glu Thr Thr Met Arg Ser Pro Val Phe Thr Asp Asn Ser Ser 210 . 215 . 220

Pro Pro Ala Val Pro Gln Ser Phe Gln Val Ala His Leu His Ala Pro 225 \$230\$ 235 240

Gly Tyr Lys Val Leu Val Leu Asn Pro Ser Val Ala Ala Thr Leu Gly 260 265 270

Phe Gly Ala Tyr Met Ser Lys Ala His Gly Ile Asp Pro Asn Ile Arg 275 280 285

Thr Gly Val Arg Thr Ile Thr Thr Gly Ala Pro Val Thr Tyr Ser Thr 290 295 300

Tyr Gly Lys Phe Leu Ala Asp Gly Gly Cys Ser Gly Gly Ala Tyr Asp 305 310315315

Ile Ile Ile Cys Asp Glu Cys His Ser Thr Asp Ser Thr Thr Ile Leu 325 330 335

Gly Ile Gly Thr Val Leu Asp Gln Ala Glu Thr Ala Gly Ala Arg Leu 340 345 350

Pro Asn Ile Glu Glu Val Ala Leu Ser Asn Thr Gly Glu Ile Pro Phe 370 375 380

Tyr Gly Lys Ala Ile Pro Ile Glu Ala Ile Arg Gly Gly Arg His Leu 385 390 395 400

Ile Phe Cys His Ser Lys Lys Lys Cys Asp Glu Leu Ala Ala Lys Leu 405 410 415

Ser Val Ile Pro Thr Ile Gly Asp Val Val Val Val Ala Thr Asp Ala 435 440 445

Thr Cys Val Thr Gln Thr Val Asp Phe Ser Leu Asp Pro Thr Phe Thr 465 \$470\$

Ile Glu Thr Thr Thr Val Pro Gln Asp Ala Val Ser Arg Ser Gln Arg 485 490 495

Arg Gly Arg Thr Gly Arg Gly Arg Gly Ile Tyr Arg Phe Val Thr

Pro Gly Glu Arg Pro Ser Gly Met Phe Asp Ser Ser Val Leu Cys Glu 515 520 525

Cys Tyr Asp Ala Gly Cys Ala Trp Tyr Glu Leu Thr Pro Ala Glu Thr

530 535 540

Ser Val Arg Leu Arg Ala Tyr Leu Asn Thr Pro Gly Leu Pro Val Cys 545 550 550 550

Gln Asp His Leu Glu Phe Trp Glu Ser Val Phe Thr Gly Leu Thr His

565 570 575

Ile Asp Ala His Phe Leu Ser Gln Thr Lys Gln Ala Gly Asp Asn Phe $580 \hspace{1.5cm} 585 \hspace{1.5cm} 585 \hspace{1.5cm} 590 \hspace{1.5cm}$

Pro Tyr Leu Val Ala Tyr Gln Ala Thr Val Cys Ala Arg Ala Gln Ala 595 600 605

Pro Pro Pro Ser Trp Asp Gln Met Trp Lys Cys Leu Ile Arg Leu Lys 610 615 620

Pro Thr Leu His Gly Pro Thr Pro Leu Leu Tyr Arg Leu Gly Ala Val 625 630 635 640

Gln Asn Glu Val Thr Leu Thr His Pro Ile Thr Lys Tyr Ile Met Ala 645 650 655

Cys Met Ser Ala Asp Leu Glu Val Val 660 665

(2) INFORMATION FOR SEO ID NO:14:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 665 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEO ID NO:14:

Met Gly Ser Ser His His His His His Ser Ser Gly Leu Val Pro

Arg Gly Ser His Met Gly Ser Val Val Ile Val Gly Arg Ile Ile Leu $20 \hspace{1cm} 25 \hspace{1cm} 30 \hspace{1cm}$

Ser Gly Ser Gly Ser Ile Thr Ala Tyr Ser Gln Gln Thr Arg Gly Leu $35 \hspace{1.5cm} 40 \hspace{1.5cm} 45 \hspace{1.5cm}$

Leu Gly Cys Lys Lys Thr Ser Leu Thr Gly Arg Asp Lys Asn Gln Val 50 60

Glu Gly Glu Val Gln Val Val Ser Thr Ala Thr Gln Ser Phe Leu Ala 65 707575

Thr Cys Val Asn Gly Val Cys Trp Thr Val Tyr His Gly Ala Gly Ser $85 \hspace{1.5cm} 90 \hspace{1.5cm} 95$

Lys Thr Leu Ala Gly Pro Lys Gly Pro Ile Thr Gln Met Tyr Thr Asn

			100					105					110		
Val	Asp	Gln 115	Asp	Leu	Val	Gly	Trp 120	Gln	Ala	Pro	Pro	Gly 125		Arg	Ser
Leu	Thr 130	Pro	Cys	Thr	Cys	Gly 135	Ser	Ser	Asp	Leu	Tyr 140		Val	Thr	Arg
His 145	Ala	Asp	Val	Ile	Pro 150	Val	Arg	Arg	Arg	Gly 155	Asp	Ser	Arg	Gly	Ser 160
Leu	Leu	Ser	Pro	Arg 165	Pro	Val	Ser	Tyr	Leu 170	Lys	Gly	Ser	Ser	Gly 175	Gly
Pro	Leu	Leu	Cys 180	Pro	Ser	Gly	His	Ala 185	Val	Gly	Ile	Phe	Arg 190	Ala	Ala
Val	Cys	Thr 195	Arg	Gly	Va1	Ala	Lys 200	Ala	Val	Asp	Phe	Val 205	Pro	Val	Glu
Ser	Met 210	Glu	Thr	Thr	Met	Arg 215	Ser	Pro	Val	Phe	Thr 220	Asp	Asn	Ser	Ser
Pro 225	Pro	Ala	Val	Pro	Gln 230	Ser	Phe	Gln	Val	Ala 235	His	Leu	His	Ala	Pro 240
Thr	Gly	Ser	Gly	Lys 245	Ser	Thr	Lys	Val	Pro 250	Ala	Ala	Tyr	Ala	Ala 255	Gln
G1y	Tyr	Lys	Va1 260	Leu	Val	Leu	Asn	Pro 265	Ser	Val	Ala	Ala	Thr 270	Leu	Gly
Phe	Gly	Ala 275	Tyr	Met	Ser	Lys	Ala 280	His	Gly	Ile	Asp	Pro 285	Asn	Ile	Arg

Thr Gly Val Arg Thr Ile Thr Thr Gly Ala Pro Val Thr Tyr Ser Thr Tyr Gly Lys Phe Leu Ala Asp Gly Gly Cys Ser Gly Gly Ala Tyr Asp Ile Ile Cys Asp Glu Cys His Ser Thr Asp Ser Thr Thr Ile Leu

Gly Ile Gly Thr Val Leu Asp Gln Ala Glu Thr Ala Gly Ala Arg Leu

Val Val Leu Ala Thr Ala Thr Pro Pro Gly Ser Val Thr Val Pro His

Pro Asn Ile Glu Glu Val Ala Leu Ser Asn Thr Gly Glu Ile Pro Phe

Tyr Gly Lys Ala Ile Pro Ile Glu Ala Ile Arg Gly Gly Arg His Leu

Ile Phe Cys His Ser Lys Lys Cys Asp Glu Leu Ala Ala Lys Leu

Ser Val Ile Pro Thr Ile Gly Asp Val Val Val Val Ala Thr Asp Ala $435 \hspace{1.5cm} 440 \hspace{1.5cm} 440 \hspace{1.5cm} 445$

Leu Met Thr Gly Tyr Thr Gly Asp Phe Asp Ser Val Ile Asp Cys Asn 450 455 460

Thr Cys Val Thr Gln Thr Val Asp Phe Ser Leu Asp Pro Thr Phe Thr 465 470 475 480

Ile Glu Thr Thr Val Pro Gln Asp Ala Val Ser Arg Ser Gln Arg
485 490 495

Pro Gly Glu Arg Pro Ser Gly Met Phe Asp Ser Ser Val Leu Cys Glu 515 520 525

Cys Tyr Asp Ala Gly Cys Ala Trp Tyr Glu Leu Thr Pro Ala Glu Thr 530 535 540

Ser Val Arg Leu Arg Ala Tyr Leu Asn Thr Pro Gly Leu Pro Val Cys 545 550 555 560

Gln Asp His Leu Glu Phe Trp Glu Ser Val Phe Thr Gly Leu Thr His $565 \hspace{1.5cm} 570 \hspace{1.5cm} 575 \hspace{1.5cm}$

Ile Asp Ala His Phe Leu Ser Gln Thr Lys Gln Ala Gly Asp Asn Phe 580 585 595

Pro Tyr Leu Val Ala Tyr Gln Ala Thr Val Cys Ala Arg Ala Gln Ala 595 $\,$ 600 $\,$ 605

Pro Pro Pro Ser Trp Asp Gln Met Trp Lys Cys Leu Ile Arg Leu Lys 610 615 620

Pro Thr Leu His Gly Pro Thr Pro Leu Leu Tyr Arg Leu Gly Ala Val 625 630 635 640

Gln Asn Glu Val Thr Leu Thr His Pro Ile Thr Lys Tyr Ile Met Ala 645 650 655

Cys Met Ser Ala Asp Leu Glu Val Val 660 665

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 665 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

Met Gly Ser Ser His His His His His Ser Ser Gly Leu Val Pro $\frac{1}{5}$ 10 15

Arg Gly Ser His Met Gly Ser Val Val Ile Val Gly Arg Ile Ile Leu $20 \hspace{1.5cm} 25 \hspace{1.5cm} 30 \hspace{1.5cm}$

Ser Gly Ser Gly Ser Ile Thr Ala Tyr Ser Gln Gln Thr Arg Gly Leu $35\,$

Leu Gly Cys Ile Ile Thr Ser Leu Thr Gly Arg Asp Lys Asn Gln Val $50 \\ 60$

Glu Gly Glu Val Gln Val Val Ser Thr Ala Thr Gln Ser Phe Leu Ala 65 70 75 80

Thr Cys Val Asn Gly Val Cys Trp Thr Val Tyr His Gly Ala Gly Ser 85 90 95

Lys Thr Leu Ala Gly Pro Lys Gly Pro Ile Thr Gln Met Tyr Thr Asn 100 105 110

Val Asp Gln Asp Leu Val Gly Trp Gln Ala Pro Pro Gly Ala Arg Ser $115 \\ 125 \\ 125$

Leu Thr Pro Cys Thr Cys Gly Ser Ser Asp Leu Tyr Leu Val Thr Arg 130 135 140

His Ala Asp Val Ile Pro Val Arg Arg Arg Gly Asp Ser Arg Gly Ser 145 \$150\$

Leu Leu Ser Pro Arg Pro Val Ser Tyr Leu Lys Gly Ser Ala Gly Gly 165 \$170\$

Pro Leu Cys Pro Ser Gly His Ala Val Gly Ile Phe Arg Ala Ala 180 \$185\$

Val Cys Thr Arg Gly Val Ala Lys Ala Val Asp Phe Val Pro Val Glu 195 200 205

Ser Met Glu Thr Thr Met Arg Ser Pro Val Phe Thr Asp Asn Ser Ser 210 215 220

Pro Pro Ala Val Pro Gln Ser Phe Gln Val Ala His Leu His Ala Pro 225 230230235

Thr Gly Ser Gly Lys Ser Thr Lys Val Pro Ala Ala Tyr Ala Ala Gln
245 250 250

Gly Tyr Lys Val Leu Val Leu Asn Pro Ser Val Ala Ala Thr Leu Gly 260 265 270

Phe Gly Ala Tyr Met Ser Lys Ala His Gly Ile Asp Pro Asn Ile Arg $275 \\ 280 \\ 285$

- Thr Gly Val Arg Thr Ile Thr Thr Gly Ala Pro Val Thr Tyr Ser Thr 290 295 300
- Tyr Gly Lys Phe Leu Ala Asp Gly Gly Cys Ser Gly Gly Ala Tyr Asp 305 \$310\$
- Ile Ile Cys Asp Glu Cys His Ser Thr Asp Ser Thr Thr Ile Leu 325 330 335
- Gly Ile Gly Thr Val Leu Asp Gln Ala Glu Thr Ala Gly Ala Arg Leu \$340\$
- Pro Asn Ile Glu Glu Val Ala Leu Ser Asn Thr Gly Glu Ile Pro Phe 370 375 380
- Tyr Gly Lys Ala Ile Pro Ile Glu Ala Ile Arg Gly Gly Arg His Leu 385 \$390\$
- Ile Phe Cys His Ser Lys Lys Lys Cys Asp Glu Leu Ala Ala Lys Leu 405 410 415
- Ser Gly Leu Gly Ile Asn Ala Val Ala Tyr Tyr Arg Gly Leu Asp Val \$420\$ \$425\$
- Ser Val Ile Pro Thr Ile Gly Asp Val Val Val Val Ala Thr Asp Ala $435 \hspace{1.5cm} 440 \hspace{1.5cm} 445 \hspace{1.5cm}$
- Leu Met Thr Gly Tyr Thr Gly Asp Phe Asp Ser Val Ile Asp Cys Asn 450 455 460
- Thr Cys Val Thr Gln Thr Val Asp Phe Ser Leu Asp Pro Thr Phe Thr 465 470 475 480
- Ile Glu Thr Thr Thr Val Pro Gln Asp Ala Val Ser Arg Ser Gln Arg
 485 490 495
- Arg Gly Arg Thr Gly Arg Gly Arg Gly Ile Tyr Arg Phe Val Thr $500 \hspace{1.5cm} 505 \hspace{1.5cm} 510 \hspace{1.5cm}$
- Pro Gly Glu Arg Pro Ser Gly Met Phe Asp Ser Ser Val Leu Cys Glu 515 $$ 520 $$ 525
- Cys Tyr Asp Ala Gly Cys Ala Trp Tyr Glu Leu Thr Pro Ala Glu Thr 530 540
- Ser Val Arg Leu Arg Ala Tyr Leu Asn Thr Pro Gly Leu Pro Val Cys 545 550 555 555
- Gln Asp His Leu Glu Phe Trp Glu Ser Val Phe Thr Gly Leu Thr His 565 570 575
- Ile Asp Ala His Phe Leu Ser Gln Thr Lys Gln Ala Gly Asp Asn Phe $580 \hspace{1.5cm} 590 \hspace{1.5cm} 590 \hspace{1.5cm}$
- Pro Tyr Leu Val Ala Tyr Gln Ala Thr Val Cys Ala Arg Ala Gln Ala

Pro Pro Pro Ser Trp Asp Gln Met Trp Lys Cys Leu Ile Arg Leu Lys 615

Pro Thr Leu His Gly Pro Thr Pro Leu Leu Tyr Arg Leu Gly Ala Val 630

Gln Asn Glu Val Thr Leu Thr His Pro Ile Thr Lys Tyr Ile Met Ala 650

Cys Met Ser Ala Asp Leu Glu Val Val 660 665

(2) INFORMATION FOR SEQ ID NO:16:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 665 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

Met Gly Ser Ser His His His His His Ser Ser Gly Leu Val Pro

Arg Gly Ser His Met Gly Ser Val Val Ile Val Gly Arg Ile Ile Leu

Ser Gly Ser Gly Ser Ile Thr Ala Tyr Ser Gln Gln Thr Arg Gly Leu

Leu Gly Cys Lys Ile Thr Ser Leu Thr Gly Arg Asp Lys Asn Gln Val

Glu Gly Glu Val Gln Val Val Ser Thr Ala Thr Gln Ser Phe Leu Ala 65 70

Thr Cys Val Asn Gly Val Cys Trp Thr Val Tyr His Gly Ala Gly Ser 90

Lys Thr Leu Ala Gly Pro Lys Gly Pro Ile Thr Gln Met Tyr Thr Asn

Val Asp Gln Asp Leu Val Gly Trp Gln Ala Pro Pro Gly Ala Arg Ser

Leu Thr Pro Cys Thr Cys Gly Ser Ser Asp Leu Tyr Leu Val Thr Arg 135 140

His Ala Asp Val Ile Pro Val Arg Arg Gly Asp Ser Arg Gly Ser 145

Leu Leu Ser Pro Arg Pro Val Ser Tyr Leu Lys Gly Ser Ala Gly Gly

175

Pro Leu Leu Cys Pro Ser Gly His Ala Val Gly Ile Phe Arg Ala Ala 180 185 190

Val Cys Thr Arg Gly Val Ala Lys Ala Val Asp Phe Val Pro Val Glu 195 200 205

Ser Met Glu Thr Thr Met Arg Ser Pro Val Phe Thr Asp Asn Ser Ser 210 215 220

Pro Pro Ala Val Pro Gln Ser Phe Gln Val Ala His Leu His Ala Pro 225 \$230\$

Thr Gly Ser Gly Lys Ser Thr Lys Val Pro Ala Ala Tyr Ala Ala Gln 245 250 255

Gly Tyr Lys Val Leu Val Leu Asn Pro Ser Val Ala Ala Thr Leu Gly $260 \hspace{1cm} 265 \hspace{1cm} 270 \hspace{1cm} .$

Phe Gly Ala Tyr Met Ser Lys Ala His Gly Ile Asp Pro Asn Ile Arg 275 280 285

Thr Gly Val Arg Thr Ile Thr Thr Gly Ala Pro Val Thr Tyr Ser Thr $290 \\ \hspace{1.5cm} 295 \\ \hspace{1.5cm} 300 \\ \hspace{1.5cm}$

Tyr Gly Lys Phe Leu Ala Asp Gly Gly Cys Ser Gly Gly Ala Tyr Asp 305 310315315315

Ile Ile Ile Cys Asp Glu Cys His Ser Thr Asp Ser Thr Thr Ile Leu 325 330 335

Gly Ile Gly Thr Val Leu Asp Gln Ala Glu Thr Ala Gly Ala Arg Leu $340 \hspace{1.5cm} 345 \hspace{1.5cm} 350$

Val Val Leu Ala Thr Ala Thr Pro Pro Gly Ser Val Thr Val Pro His 355 360 365

Pro Asn Ile Glu Glu Val Ala Leu Ser Asn Thr Gly Glu Ile Pro Phe 370 385

Tyr Gly Lys Ala Ile Pro Ile Glu Ala Ile Arg Gly Gly Arg His Leu 385 390 395 400

Ile Phe Cys His Ser Lys Lys Lys Cys Asp Glu Leu Ala Ala Lys Leu 405 410 415

Ser Gly Leu Gly Ile Asn Ala Val Ala Tyr Tyr Arg Gly Leu Asp Val $420 \hspace{1.5cm} 425 \hspace{1.5cm} 430$

Ser Val Ile Pro Thr Ile Gly Asp Val Val Val Val Ala Thr Asp Ala 435 440 445

Leu Met Thr Gly Tyr Thr Gly Asp Phe Asp Ser Val Ile Asp Cys Asn 450 455 460

Thr Cys Val Thr Gln Thr Val Asp Phe Ser Leu Asp Pro Thr Phe Thr 465 470 475 480

Ile Glu Thr Thr Thr Val Pro Gln Asp Ala Val Ser Arg Ser Gln Arg 485 490 495

Pro Gly Glu Arg Pro Ser Gly Met Phe Asp Ser Ser Val Leu Cys Glu 515 520 525

Cys Tyr Asp Ala Gly Cys Ala Trp Tyr Glu Leu Thr Pro Ala Glu Thr 530 535 540

Ser Val Arg Leu Arg Ala Tyr Leu Asn Thr Pro Gly Leu Pro Val Cys 545 550 550 555 560

Gln Asp His Leu Glu Phe Trp Glu Ser Val Phe Thr Gly Leu Thr His 565 570 575

Ile Asp Ala His Phe Leu Ser Gln Thr Lys Gln Ala Gly Asp Asn Phe 580 585 590

Pro Tyr Leu Val Ala Tyr Gln Ala Thr Val Cys Ala Arg Ala Gln Ala 595 600 605

Pro Pro Pro Ser Trp Asp Gln Met Trp Lys Cys Leu Ile Arg Leu Lys 610 615 620

Pro Thr Leu His Gly Pro Thr Pro Leu Leu Tyr Arg Leu Gly Ala Val 625 $$ 630 $$ 635 $$ 640

Gln Asn Glu Val Thr Leu Thr His Pro Ile Thr Lys Tyr Ile Met Ala $645 \hspace{1.5cm} 650 \hspace{1.5cm} 655$

Cys Met Ser Ala Asp Leu Glu Val Val 660 665

(2) INFORMATION FOR SEO ID NO:17:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 665 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

Met Gly Ser Ser His His His His His Ser Ser Gly Leu Val Pro $1 \hspace{1cm} 5 \hspace{1cm} 10 \hspace{1cm} 15$

Arg Gly Ser His Met Gly Ser Val Val Ile Val Gly Arg Ile Ile Leu 20 25 30

Ser Gly Ser Gly Ser Ile Thr Ala Tyr Ser Gln Gln Thr Arg Gly Leu $35 \hspace{1.5cm} 40 \hspace{1.5cm} 45$

- Glu Gly Glu Val Gln Val Val Ser Thr Ala Thr Gln Ser Phe Leu Ala $65 \hspace{1.5cm} 70 \hspace{1.5cm} 75 \hspace{1.5cm} 80$
- Thr Cys Val Asn Gly Val Cys Trp Thr Val Tyr His Gly Ala Gly Ser $85 \hspace{1.5cm} 90 \hspace{1.5cm} 95$
- Lys Thr Leu Ala Gly Pro Lys Gly Pro Ile Thr Gln Met Tyr Thr Asn $100 \\ 105 \\ 110$
- Val Asp Gln Asp Leu Val Gly Trp Gln Ala Pro Pro Gly Ala Arg Ser $^\circ$ 115 \$120 \$125
- Leu Thr Pro Cys Thr Cys Gly Ser Ser Asp Leu Tyr Leu Val Thr Arg 130 135 140
- His Ala Asp Val Ile Pro Val Arg Arg Arg Gly Asp Ser Arg Gly Ser 145
- Leu Leu Ser Pro Arg Pro Val Ser Tyr Leu Lys Gly Ser Ala Gly Gly 165 170 175
- Pro Leu Cys Pro Ser Gly His Ala Val Gly Ile Phe Arg Ala Ala 180 185 190
- Val Cys Thr Arg Gly Val Ala Lys Ala Val Asp Phe Val Pro Val Glu 195 200 205
- Ser Met Glu Thr Thr Met Arg Ser Pro Val Phe Thr Asp Asn Ser Ser 210 215 220
- Pro Pro Ala Val Pro Gln Ser Phe Gln Val Ala His Leu His Ala Pro 225 $$ 230 $$ 235 $$ 240
- Thr Gly Ser Gly Lys Ser Thr Lys Val Pro Ala Ala Tyr Ala Ala Gln
- Gly Tyr Lys Val Leu Val Leu Asn Pro Ser Val Ala Ala Thr Leu Gly $260 \hspace{1cm} 265 \hspace{1cm} 265 \hspace{1cm} 270 \hspace{1cm}$
- Phe Gly Ala Tyr Met Ser Lys Ala His Gly Ile Asp Pro Asn Ile Arg 275 280 285
- Thr Gly Val Arg Thr Ile Thr Thr Gly Ala Pro Val Thr Tyr Ser Thr 290 \$295 \$300
- Tyr Gly Lys Phe Leu Ala Asp Gly Gly Cys Ser Gly Gly Ala Tyr Asp 305 \$310\$
- Ile Ile Ile Cys Asp Glu Cys His Ser Thr Asp Ser Thr Thr Ile Leu \$325\$ \$330\$ \$335
- Gly Ile Gly Thr Val Leu Asp Gln Ala Glu Thr Ala Gly Ala Arg Leu $340 \hspace{1.5cm} 345 \hspace{1.5cm} 350 \hspace{1.5cm}$

- Val Val Leu Ala Thr Ala Thr Pro Pro Gly Ser Val Thr Val Pro His 355 360 365
- Pro Asn Ile Glu Glu Val Ala Leu Ser Asn Thr Gly Glu Ile Pro Phe 370 380
- Tyr Gly Lys Ala Ile Pro Ile Glu Ala Ile Arg Gly Gly Arg His Leu 385 390 395 400
- Ile Phe Cys His Ser Lys Lys Lys Cys Asp Glu Leu Ala Ala Lys Leu 405 410 415
- Ser Gly Leu Gly Ile Asn Ala Val Ala Tyr Tyr Arg Gly Leu Asp Val \$420\$.
- Ser Val Ile Pro Thr Ile Gly Asp Val Val Val Val Ala Thr Asp Ala $435 \hspace{1.5cm} 440 \hspace{1.5cm} 445$
- Leu Met Thr Gly Tyr Thr Gly Asp Phe Asp Ser Val Ile Asp Cys Asn 450 455 460
- Thr Cys Val Thr Gln Thr Val Asp Phe Ser Leu Asp Pro Thr Phe Thr 465 470470475
- Arg Gly Arg Thr Gly Arg Gly Arg Gly Ile Tyr Arg Phe Val Thr 500 505 510
- Pro Gly Glu Arg Pro Ser Gly Met Phe Asp Ser Ser Val Leu Cys Glu 515 $$ 520 $$ 525
- Ser Val Arg Leu Arg Ala Tyr Leu Asn Thr Pro Gly Leu Pro Val Cys 545 555 560
- Gln Asp His Leu Glu Phe Trp Glu Ser Val Phe Thr Gly Leu Thr His $565 \hspace{1.5cm} 575 \hspace{1.5cm} 575 \hspace{1.5cm}$
- Ile Asp Ala His Phe Leu Ser Gln Thr Lys Gln Ala Gly Asp Asn Phe 580 590
- Pro Tyr Leu Val Ala Tyr Gln Ala Thr Val Cys Ala Arg Ala Gln Ala 595 $\,$ 600 $\,$ 605
- Pro Pro Pro Ser Trp Asp Gln Met Trp Lys Cys Leu Ile Arg Leu Lys 610 615 620
- Pro Thr Leu His Gly Pro Thr Pro Leu Leu Tyr Arg Leu Gly Ala Val 625 630 635 640
- Gln Asn Glu Val Thr Leu Thr His Pro Ile Thr Lys Tyr Ile Met Ala $645 \hspace{1cm} 655 \hspace{1cm} 655$
- Cys Met Ser Ala Asp Leu Glu Val Val

- (2) INFORMATION FOR SEO ID NO:18:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 665 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: protein
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

Met Gly Ser Ser His His His His His Ser Ser Gly Leu Val Pro

Arg Gly Ser His Met Gly Ser Val Val Ile Val Gly Arg Ile Ile Leu. 25

Ser Gly Ser Gly Ser Ile Thr Ala Tyr Ser Gln Gln Thr Arg Gly Leu

Leu Gly Cys Lys Lys Thr Ser Leu Thr Gly Arg Asp Lys Asn Gln Val

Glu Gly Glu Val Gln Val Val Ser Thr Ala Thr Gln Ser Phe Leu Ala

Thr Cys Val Asn Gly Val Cys Trp Thr Val Tyr His Gly Ala Gly Ser

Lys Thr Leu Ala Gly Pro Lys Gly Pro Ile Thr Gln Met Tyr Thr Asn

Val Asp Gln Asp Leu Val Gly Trp Gln Ala Pro Pro Gly Ala Arg Ser 120

Leu Thr Pro Cys Thr Cys Gly Ser Ser Asp Leu Tyr Leu Val Thr Arg 130

His Ala Asp Val Ile Pro Val Arg Arg Gly Asp Ser Arg Gly Ser 150

Leu Leu Ser Pro Arg Pro Val Ser Tyr Leu Lys Gly Ser Ala Gly Gly 165 170

Pro Leu Cys Pro Ser Gly His Ala Val Gly Ile Phe Arg Ala Ala

Val Cys Thr Arg Gly Val Ala Lys Ala Val Asp Phe Val Pro Val Glu

Ser Met Glu Thr Thr Met Arg Ser Pro Val Phe Thr Asp Asn Ser Ser 210 215

Pro Pro Ala Val Pro Gln Ser Phe Gln Val Ala His Leu His Ala Pro

225					230					235					240
Thr	Gly	Ser	Gly	Lys 245	Ser	Thr	Lys	Va1	Pro 250	Ala	Ala	Tyr	Ala	Ala 255	Gln
Gly	Tyr	Lys	Val 260	Leu	Val	Leu	Asn	Pro 265		Val	Ala	Ala	Thr 270		Gly
Phe	Gly	Ala 275	Tyr	Met	Ser	Lys	Ala 280	His	Gly	Ile	Asp	Pro 285		Ile	Arg
Thr	Gly 290	Val	Arg	Thr	Ile	Thr 295	Thr	Gly	Ala	Pro	Val 300		Tyr	Ser	Thr
Tyr 305	Gly	Lys	Phe	Leu	Ala 310	Asp	Gly	Gly	Cys	Ser 315	Gly	Gly	Ala	Tyr	Asp 320
Ile	Ile	Ile	Cys	Asp 325	Glu	Cys	His	Ser	Thr 330	Asp	Ser	Thr	Thr	Ile 335	Leu
Gly	Ile	Gly	Thr 340	Val	Leu	Asp	Gln	Ala 345	Glu	Thr	Ala	Gly	Ala 350	Arg	Leu
Val	Val	Leu 355	Ala	Thr	Ala	Thr	Pro 360	Pro	Gly	Ser	Val	Thr 365	Val	Pro	His
Pro	Asn 370	Ile	Glu	Glu	Va1	Ala 375	Leu	Ser	Asn	Thr	Gly 380	Glu	Ile	Pro	Phe
Tyr 385	Gly	Lys	Ala	Ile	Pro 390	Ile	Glu	Ala	Ile	Arg 395	Gly	Gly	Arg	His	Leu 400
Ile	Phe	Cys	His	Ser 405	Lys	Lys	Lys	Cys	Asp 410	Glu	Leu	Ala	Ala	Lys 415	Leu
Ser	Gly	Leu	Gly 420	Ile	Asn	Ala	Val	Ala 425	Tyr	Tyr	Arg	Gly	Leu 430	Asp	Val
Ser	Val	Ile 435	Pro	Thr	Ile	Gly	Asp 440	Val	Val	Val	Val	Ala 445	Thr	Asp	Ala
Leu	Met 450	Thr	Gly	Tyr	Thr	Gly 455	Asp	Phe	Asp	Ser	Val 460	Ile	Asp	Cys	Asn
Thr 465	Cys	Val	Thr	Gln	Thr 470	Val	Asp	Phe	Ser	Leu 475	Asp	Pro	Thr	Phe	Thr 480
Ile	Glu	Thr	Thr	Thr 485	Val	Pro	Gln	Asp	Ala 490	Val	Ser	Arg	Ser	Gln 495	Arg
Arg	G1y	Arg	Thr 500	Gly	Arg	Gly	Arg	Arg 505	Gly	Ile	Tyr	Arg	Phe 510	Val	Thr
Pro	Gly	Glu 515	Arg	Pro	Ser	Gly	Met 520	Phe	Asp	Ser	Ser	Val 525	Leu	Cys	Glu
Cys	Tyr 530	Asp	Ala	Gly	Cys	Ala 535	Trp	Tyr	Glu	Leu	Thr 540	Pro	Ala	Glu	Thr

Ser Val Arg Leu Arg Ala Tyr Leu Asn Thr Pro Gly Leu Pro Val Cys 545 555 560

Gln Asp His Leu Glu Phe Trp Glu Ser Val Phe Thr Gly Leu Thr His 565 570 575

Ile Asp Ala His Phe Leu Ser Gln Thr Lys Gln Ala Gly Asp Asn Phe 580 585 590

Pro Tyr Leu Val Ala Tyr Gln Ala Thr Val Cys Ala Arg Ala Gln Ala $595 \hspace{1.5cm} 600 \hspace{1.5cm} 605 \hspace{1.5cm}$

Pro Pro Pro Ser Trp Asp Gln Met Trp Lys Cys Leu Ile Arg Leu Lys 610 615 620

Pro Thr Leu His Gly Pro Thr Pro Leu Leu Tyr Arg Leu Gly Ala Val 625 $$ 630 $$ 635 $$ 640

Gln Asn Glu Val Thr Leu Thr His Pro Ile Thr Lys Tyr Ile Met Ala $645 \hspace{1cm} 655$

Cys Met Ser Ala Asp Leu Glu Val Val 660 665

- (2) INFORMATION FOR SEO ID NO:19:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 671 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: protein
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

Met Gly Ser Ser His His His His His Ber Ser Gly Leu Val Pro 1 5 10 10 15

Arg Gly Ser His Met Ala Tyr Ser Leu Thr Thr Gly Ser Val Val Ile 20 25 30

Val Gly Arg Ile Ile Leu Ser Gly Ser Gly Ser Ile Thr Ala Tyr Ser 35 40 45

Gln Gln Thr Arg Gly Leu Leu Gly Cys Ile Ile Thr Ser Leu Thr Gly 50 60

Arg Asp Lys Asn Gln Val Glu Gly Glu Val Gln Val Val Ser Thr Ala 65 70 75 80

Thr Gln Ser Phe Leu Ala Thr Cys Val Asn Gly Val Cys Trp Thr Val $85 \\ 90 \\ 95$

Tyr His Gly Ala Gly Ser Lys Thr Leu Ala Gly Pro Lys Gly Pro Ile $100 \hspace{1cm} 105 \hspace{1cm} 110 \hspace{1cm}$

- Thr Gln Met Tyr Thr Asn Val Asp Gln Asp Leu Val Gly Trp Gln Ala 115 120 125
- Pro Pro Gly Ala Arg Ser Leu Thr Pro Cys Thr Cys Gly Ser Ser Asp 130 135
- Leu Tyr Leu Val Thr Arg His Ala Asp Val Ile Pro Val Arg Arg Arg 145 \$150\$
- Gly Asp Ser Arg Gly Ser Leu Leu Ser Pro Arg Pro Val Ser Tyr Leu 165 170 175
- Lys Gly Ser Ser Gly Gly Pro Leu Leu Cys Pro Ser Gly His Ala Val 180 185 190
- Gly Ile Phe Arg Ala Ala Val Cys Thr Arg Gly Val Ala Lys Ala Val 195 200200205
- Asp Phe Val Pro Val Glu Ser Met Glu Thr Thr Met Arg Ser Pro Val 210 215 220
- Phe Thr Asp Asn Ser Ser Pro Pro Ala Val Pro Gln Ser Phe Gln Val 225 230 235
- Ala His Leu His Ala Pro Thr Gly Ser Gly Lys Ser Thr Lys Val Pro $245 \hspace{1cm} 250 \hspace{1cm} 250 \hspace{1cm} 255 \hspace{1cm}$
- Ala Ala Tyr Ala Ala Gln Gly Tyr Lys Val Leu Val Leu Asn Pro Ser $260 \hspace{1.5cm} 265 \hspace{1.5cm} 270 \hspace{1.5cm}$
- Val Ala Ala Thr Leu Gly Phe Gly Ala Tyr Met Ser Lys Ala His Gly $275 \\ 280 \\ 285$
- Ile Asp Pro Asn Ile Arg Thr Gly Val Arg Thr Ile Thr Thr Gly Ala $290 \hspace{1.5cm} 295 \hspace{1.5cm} 300 \hspace{1.5cm}$
- Pro Val Thr Tyr Ser Thr Tyr Gly Lys Phe Leu Ala Asp Gly Gly Cys $305 \hspace{1.5cm} 310 \hspace{1.5cm} 315 \hspace{1.5cm} 320 \hspace{1.5cm}$
- Ser Gly Gly Ala Tyr Asp Ile Ile Cys Asp Glu Cys His Ser Thr \$325\$
- Asp Ser Thr Thr Ile Leu Gly Ile Gly Thr Val Leu Asp Gln Ala Glu $340 \hspace{1.5cm} 345 \hspace{1.5cm} 350 \hspace{1.5cm}$
- Thr Ala Gly Ala Arg Leu Val Val Leu Ala Thr Ala Thr Pro Pro Gly 355 360 365
- Ser Val Thr Val Pro His Pro Asn Ile Glu Glu Val Ala Leu Ser Asn $370 \hspace{1.5cm} 375 \hspace{1.5cm} 380 \hspace{1.5cm}$
- Thr Gly Glu Ile Pro Phe Tyr Gly Lys Ala Ile Pro Ile Glu Ala Ile 385 390 395 400
- Arg Gly Gly Arg His Leu Ile Phe Cys His Ser Lys Lys Lys Cys Asp \$405\$

- Glu Leu Ala Ala Lys Leu Ser Gly Leu Gly Ile Asn Ala Val Ala Tyr $420 \hspace{1.5cm} 425 \hspace{1.5cm} 430 \hspace{1.5cm}$
- Tyr Arg Gly Leu Asp Val Ser Val Ile Pro Thr Ile Gly Asp Val Val 435 445
- Val Val Ala Thr Asp Ala Leu Met Thr Gly Tyr Thr Gly Asp Phe Asp 450 455
- Ser Val Ile Asp Cys Asn Thr Cys Val Thr Gln Thr Val Asp Phe Ser 465 470 475
- Leu Asp Pro Thr Phe Thr Ile Glu Thr Thr Thr Val Pro Gln Asp Ala
 485 490
- Val Ser Arg Ser Gln Arg Arg Gly Arg Thr Gly Arg Gly Arg Gly 500 505 510
- Ser Ser Val Leu Cys Glu Cys Tyr Asp Ala Gly Cys Ala Trp Tyr Glu 530 535 540
- Leu Thr Pro Ala Glu Thr Ser Val Arg Leu Arg Ala Tyr Leu Asn Thr 545 550 555 560
- Pro Gly Leu Pro Val Cys Gln Asp His Leu Glu Phe Trp Glu Ser Val 565 570 575
- Phe Thr Gly Leu Thr His Ile Asp Ala His Phe Leu Ser Gln Thr Lys 580 585 590

- Cys Leu Ile Arg Leu Lys Pro Thr Leu His Gly Pro Thr Pro Leu Leu 625 630 635 640
- Thr Lys Tyr Ile Met Ala Cys Met Ser Ala Asp Leu Glu Val Val 660 665 670
- (2) INFORMATION FOR SEQ ID NO:20:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 671 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

Met Gly Ser Ser His His His His His Ser Ser Gly Leu Val Pro $1 \hspace{1cm} 5 \hspace{1cm} 10 \hspace{1cm} 15$

Arg Gly Ser His Met Ala Tyr Ser Leu Thr Thr Gly Ser Val Val Ile 20 25 30

Val Gly Arg Ile Ile Leu Ser Gly Ser Gly Ser Ile Thr Ala Tyr Ser 35 40 45

Gln Gln Thr Arg Gly Leu Leu Gly Cys Ile Ile Thr Ser Leu Thr Gly $_{50}$

Arg Asp Lys Asn Gln Val Glu Gly Glu Val Gln Val Val Ser Thr Ala $65 \hspace{1.5cm} 70 \hspace{1.5cm} 75 \hspace{1.5cm} 80$

Thr Gln Ser Phe Leu Ala Thr Cys Val Asn Gly Val Cys Trp Thr Val $85 \\ 90 \\ 95$

Thr Gln Met Tyr Thr Asn Val Asp Gln Asp Leu Val Gly Trp Gln Ala 115 120 125

Pro Pro Gly Ala Arg Ser Leu Thr Pro Cys Thr Cys Gly Ser Ser Asp $_{\rm I}\,130$

Leu Tyr Leu Val Thr Arg His Ala Asp Val Ile Pro Val Arg Arg Arg 145 150 155 160

Gly Asp Ser Arg Gly Ser Leu Leu Ser Pro Arg Pro Val Ser Tyr Leu 165 170 175

Lys Gly Ser Ala Gly Gly Pro Leu Leu Cys Pro Ser Gly His Ala Val \$180\$

Gly Ile Phe Arg Ala Ala Val Cys Thr Arg Gly Val Ala Lys Ala Val 195 200 205

Asp Phe Val Pro Val Glu Ser Met Glu Thr Thr Met Arg Ser Pro Val 210 215 220

Phe Thr Asp Asn Ser Ser Pro Pro Ala Val Pro Gln Ser Phe Gln Val 225 230 235

Ala His Leu His Ala Pro Thr Gly Ser Gly Lys Ser Thr Lys Val Pro 245 250 255

Ala Ala Tyr Ala Ala Gln Gly Tyr Lys Val Leu Val Leu Asn Pro Ser \$260\$

Val Ala Ala Thr Leu Gly Phe Gly Ala Tyr Met Ser Lys Ala His Gly 275 280 285

Ile Asp Pro Asn Ile Arg Thr Gly Val Arg Thr Ile Thr Thr Gly Ala

290 295 300 Pro Val Thr Tyr Ser Thr Tyr Gly Lys Phe Leu Ala Asp Gly Gly Cys 315 Ser Gly Gly Ala Tyr Asp Ile Ile Ile Cys Asp Glu Cys His Ser Thr Asp Ser Thr Thr Ile Leu Gly Ile Gly Thr Val Leu Asp Gln Ala Glu Thr Ala Gly Ala Arg Leu Val Val Leu Ala Thr Ala Thr Pro Pro Gly 355 360 Ser Val Thr Val Pro His Pro Asn Ile Glu Glu Val Ala Leu Ser Asn Thr Gly Glu Ile Pro Phe Tyr Gly Lys Ala Ile Pro Ile Glu Ala Ile 390 395 Arg Gly Gly Arg His Leu Ile Phe Cys His Ser Lys Lys Lys Cys Asp 405 410 Glu Leu Ala Ala Lys Leu Ser Gly Leu Gly Ile Asn Ala Val Ala Tyr Tyr Arg Gly Leu Asp Val Ser Val Ile Pro Thr Ile Gly Asp Val Val 435 440 Val Val Ala Thr Asp Ala Leu Met Thr Gly Tyr Thr Gly Asp Phe Asp Ser Val Ile Asp Cys Asn Thr Cys Val Thr Gln Thr Val Asp Phe Ser 470 475 Leu Asp Pro Thr Phe Thr Ile Glu Thr Thr Thr Val Pro Gln Asp Ala 485 Val Ser Arg Ser Gln Arg Arg Gly Arg Thr Gly Arg Gly Arg Gly 505 Ile Tyr Arg Phe Val Thr Pro Gly Glu Arg Pro Ser Gly Met Phe Asp 515 520 Ser Ser Val Leu Cys Glu Cys Tyr Asp Ala Gly Cys Ala Trp Tyr Glu Leu Thr Pro Ala Glu Thr Ser Val Arg Leu Arg Ala Tyr Leu Asn Thr 555 Pro Gly Leu Pro Val Cys Gln Asp His Leu Glu Phe Trp Glu Ser Val 565 570 Phe Thr Gly Leu Thr His Ile Asp Ala His Phe Leu Ser Gln Thr Lys

580 585 590

Gln Ala Gly Asp Asn Phe Pro Tyr Leu Val Ala Tyr Gln Ala Thr Val

600

595

Cys Leu Ile Arg Leu Lys Pro Thr Leu His Gly Pro Thr Pro Leu Leu 625 630 635 640

Tyr Arg Leu Gly Ala Val Gln Asn Glu Val Thr Leu Thr His Pro Ile 645 650 655

Thr Lys Tyr Ile Met Ala Cys Met Ser Ala Asp Leu Glu Val Val 660 665 670

- (2) INFORMATION FOR SEO ID NO:21:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 4 amino acids
 - (B) TYPE: amino acid (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: peptide
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:
 - Gly Ser Gly Ser
- (2) INFORMATION FOR SEQ ID NO:22:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 4 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: peptide

 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

Pro Ala Gly Gly

- (2) INFORMATION FOR SEQ ID NO:22:
 - - (i) SEQUENCE CHARACTERISTICS:(A) LENGTH: 1964 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: cDNA

- (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION: 1..1964
- (2) INFORMATION FOR SEQ ID NO:23:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 632 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: protein
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:
 - Met Ala Pro Ile Thr Ala Tyr Ser Gln Gln Thr Arg Gly Leu Leu Gly 1 5 10 15
 - Cys Ile Ile Thr Ser Leu Thr Gly Arg Asp Lys Asn Gln Val Glu Gly $20 \\ 25 \\ 30$
 - Glu Val Gln Val Val Ser Thr Ala Thr Gln Ser Phe Leu Ala Thr Cys 35 40 45
 - Val Asn Gly Val Cys Trp Thr Val Tyr His Gly Ala Gly Ser Lys Thr 50 55 60
 - Leu Ala Gly Pro Lys Gly Pro Ile Thr Gln Met Tyr Thr Asn Val Asp 65 70 75 80
 - Gln Asp Leu Val Gly Trp Gln Ala Pro Pro Gly Ala Arg Ser Leu Thr 85 90 95
 - Pro Cys Thr Cys Gly Ser Ser Asp Leu Tyr Leu Val Thr Arg His Ala 100 105 110
 - Asp Val Ile Pro Val Arg Arg Arg Gly Asp Ser Arg Gly Ser Leu Leu 115 120 125
 - Ser Pro Arg Pro Val Ser Tyr Leu Lys Gly Ser Ser Gly Gly Pro Leu 130 \$135\$
 - Leu Cys Pro Ser Gly His Ala Val Gly Ile Phe Arg Ala Ala Val Cys 145 150 150 155
 - Thr Arg Gly Val Ala Lys Ala Val Asp Phe Val Pro Val Glu Ser Met $165 \hspace{1cm} 170 \hspace{1cm} 175$
 - Glu Thr Thr Met Arg Ser Pro Val Phe Thr Asp Asn Ser Ser Pro Pro 180 \$185
 - Ala Val Pro Gln Ser Phe Gln Val Ala His Leu His Ala Pro Thr Gly 195 \$200\$

Ser Gly Lys Ser Thr Lys Val Pro Ala Ala Tyr Ala Ala Gln Gly Tyr 210 215

Lys Val Leu Val Leu Asn Pro Ser Val Ala Ala Thr Leu Gly Phe Gly 225 230235235235

Ala Tyr Met Ser Lys Ala His Gly Ile Asp Pro Asn Ile Arg Thr Gly 245 250 255

Val Arg Thr Ile Thr Thr Gly Ala Pro Val Thr Tyr Ser Thr Tyr Gly 260 265 270

Lys Phe Leu Ala Asp Gly Gly Cys Ser Gly Gly Ala Tyr Asp Ile Ile 275 \$280\$

Ile Cys Asp Glu Cys His Ser Thr Asp Ser Thr Thr Ile Leu Gly Ile 290 295 300

Gly Thr Val Leu Asp Gln Ala Glu Thr Ala Gly Ala Arg Leu Val Val 305 310 315 320

Leu Ala Thr Ala Thr Pro Pro Gly Ser Val Thr Val Pro His Pro Asn 325 330 335

Ile Glu Glu Val Ala Leu Ser Asn Thr Gly Glu Ile Pro Phe Tyr Gly $340 \hspace{1.5cm} 345 \hspace{1.5cm} 350 \hspace{1.5cm}$

Lys Ala Ile Pro Ile Glu Ala Ile Arg Gly Gly Arg His Leu Ile Phe $355 \hspace{1cm} 360 \hspace{1cm} 365$

Cys His Ser Lys Lys Lys Cys Asp Glu Leu Ala Ala Lys Leu Ser Gly 370 375 380

Leu Gly Ile Asn Ala Val Ala Tyr Tyr Arg Gly Leu Asp Val Ser Val 385 390 395 400

Ile Pro Thr Ile Gly Asp Val Val Val Val Ala Thr Asp Ala Leu Met \$405\$

Thr Gly Tyr Thr Gly Asp Phe Asp Ser Val Ile Asp Cys Asn Thr Cys 420 425 430

Val Thr Gln Thr Val Asp Phe Ser Leu Asp Pro Thr Phe Thr Ile Glu 435 440

Thr Thr Val Pro Gln Asp Ala Val Ser Arg Ser Gln Arg Arg Gly 450 455 460

Arg Thr Gly Arg Gly Arg Gly Ile Tyr Arg Phe Val Thr Pro Gly 465 \$470\$

Glu Arg Pro Ser Gly Met Phe Asp Ser Ser Val Leu Cys Glu Cys Tyr 485 490 495

Arg Leu Arg Ala Tyr Leu Asn Thr Pro Gly Leu Pro Val Cys Gln Asp 515 520 525

His Leu Glu Phe Trp Glu Ser Val Phe Thr Gly Leu Thr His Ile Asp 530 535 540

Ala His Phe Leu Ser Gln Thr Lys Gln Ala Gly Asp Asn Phe Pro Tyr 545 550 555

Leu Val Ala Tyr Gln Ala Thr Val Cys Ala Arg Ala Gln Ala Pro Pro 565 570 575

Pro Ser Trp Asp Gln Met Trp Lys Cys Leu Ile Arg Leu Lys Pro Thr

Leu His Gly Pro Thr Pro Leu Leu Tyr Arg Leu Gly Ala Val Gln Asn $595 \hspace{1.5cm} 600 \hspace{1.5cm} 605$

Ser Ala Asp Leu Glu Val Val Thr 625 630

(2) INFORMATION FOR SEQ ID NO:24:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 54 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

Ser Thr Trp Val Leu Val Gly Gly Val Leu Ala Ala Leu Ala Ala Tyr 1 5 10 15

Cys Leu Thr Thr Gly Ser Val Val Ile Val Gly Arg Ile Ile Leu Ser 20 25 30

Gly Arg Pro Ala Ile Val Pro Asp Arg Glu Leu Leu Tyr Gln Glu Phe $35 \hspace{1cm} 40 \hspace{1cm} 45 \hspace{1cm}$

Asp Glu Met Glu Glu Cys 50

(2) INFORMATION FOR SEQ ID NO:25:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 16 amino acids
 - (B) TYPE: amino acid

(C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(ii) MOLECULE TYPE: peptide	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:	
Asp Thr Glu Asp Val Val Cys Cys Ser Met Tyr Thr Trp Thr Gly Lys	
1 5 10 15	
(2) INFORMATION FOR SEQ ID NO:26:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 78 base pairs (B) TYPE: nucleic acid .	
(C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(ii) MOLECULE TYPE: cDNA	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:	
GATATACATA TGGGTTCTGT TGTTATTGTT GGTAGAATTA TTTTATCTGG TAGTGGTAGT	60
ATCACGGCCT ACTCCCAA	78
(2) INFORMATION FOR SEQ ID NO:27:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 36 base pairs (B) TYPE: nucleic acid	
(C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(ii) MOLECULE TYPE: CDNA	
. ,	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:	
CTCAGCGAAT TCTCAAGACC GCATAGTAGT TTCCAT	36
(2) INFORMATION FOR SEQ ID NO:28:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 39 base pairs (B) TYPE: nucleic acid	
(C) STRANDEDNESS: single (D) TOPOLOGY: linear	

(ii) MOLECULE TYPE: cDNA

	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:	
CGG	GGCCTAC TTGGTTGCAA GATCACTAGC CTTACAGGC	3
(2)	INFORMATION FOR SEQ ID NO:29:	
	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 39 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: cDNA	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:	
GCC	TGTAAGG CTAGTGATCT TGCAACCAAG TAGGCCCCG	3 9
(2)	INFORMATION FOR SEQ ID NO:30:	
	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 39 base pairs (B) TYPE: nucleic acid (C) STRANDENDESS: single (D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: cDNA	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:	
CGGC	GGCCTAC TTGGTTGCAT CAAGACTAGC CTTACAGGC	39
(2)	INFORMATION FOR SEQ ID NO:31:	
	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 39 base pairs (B) TYPE: nucleic acid (C) STRANDENNESS: single (D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: cDNA	

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31: GCCTGTAAGG CTAGTCTTGA TGCAACCAAG TAGGCCCCG

(2) INFORMATION FOR SEQ ID NO:32:	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 39 base pairs (B) TYPE: nucleic accid (C) STRANDEDMESS: single (D) TOPOLOGY: linear	
(ii) MOLECULE TYPE: cDNA	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:	
CGGGGCCTAC TTGGTTGCAA GAAGACTAGC CTTACAGGC	39
(2) INFORMATION FOR SEQ ID NO:33:	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 39 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(ii) MOLECULE TYPE: cDNA	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:	
GCCTGTAAGG CTAGTCTTCT TGCAACCAAG TAGGCCCCG	39
(2) INFORMATION FOR SEQ ID NO:34:	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 40 base pairs (B) TYPE: nucleic acid (C) STRANDENNESS: single (D) TOPOLOGY: linear	
(ii) MOLECULE TYPE: cDNA	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:	
CTCCTACTTG AAGGGCTCTG CTGGTGGTCC ACTGCTCTGC	40
(2) INFORMATION FOR SEQ ID NO:35:	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 40 base pairs (B) TYPE: nucleic acid (C) STRUNDENNESS: circle	

(D) TOPOLOGY: linear	
(ii) MOLECULE TYPE: cDNA	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:	
GCAGAGCAGT GGACCACCAG CAGAGCCCTT CAAGTAGGAG	40
(2) INFORMATION FOR SEQ ID NO:36:	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 39 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(ii) MOLECULE TYPE: cDNA	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:	
CGGGGCCTAC TTGGTTGCAA GATCACTAGC CTTACAGGC	39
(2) INFORMATION FOR SEQ ID NO:37:	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 39 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(ii) MOLECULE TYPE: cDNA	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:	
GCCTGTAAGG CTAGTGATCT TGCAACCAAG TAGGCCCCG	39
(2) INFORMATION FOR SEQ ID NO:38:	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 39 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(ii) MOLECULE TYPE: cDNA	

	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:	
CGG	GCCTAC TTGGTTGCAT CAAGACTAGC CTTACAGGC	39
(2)	INFORMATION FOR SEQ ID NO:39:	
	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 39 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: cDNA	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:	
GCC'	GTAAGG CTAGTCTTGA TGCAACCAAG TAGGCCCCG	39
(2)	INFORMATION FOR SEQ ID NO:40:	
	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 39 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: cDNA	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:	
CGGC	GCCTAC TTGGTTGCAA GAAGACTAGC CTTACAGGC	39
(2)	INFORMATION FOR SEQ ID NO:41:	
	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 39 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: cDNA	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:	
GCCI	GTAAGG CTAGTCTTGA TGCAACCAAG TAGGCCCCG	39
(2)	INFORMATION FOR SEQ ID NO:42:	
	(i) SEQUENCE CHARACTERISTICS:	

(A) LENGTH: 78 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(ii) MOLECULE TYPE: cDNA	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:	
GATATACATA TGGGTTCTGT TGTTATTGTT GGTAGAATTA TTTTATCTCC TGCTGGTGGT .	60
ATCACGGCCT ACTCCCAA	78
(2) INFORMATION FOR SEQ ID NO:43:	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 36 base pairs (B) TYPE: nucleic acid (C) STRANDEDMESS: single (D) TOPOLOGY: linear	
(ii) MOLECULE TYPE: cDNA	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43: CTCAGCGAAT TCTCAAGACC GCATAGTAGT TTCCAT (2) INFORMATION FOR SEQ ID NO:44: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 39 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear (ii) MOLECULE TYPE: cDNA	36
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44: CGGGGCCTAC TTGGTTGCAA GATCACTAGC CTTACAGGC (2) INFORMATION FOR SEQ ID NO:45: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 39 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single	39
(D) TOPOLOGY: linear	

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

GCCTGTAAGG CTAGTGATCT TGCAACCAAG TAGGCCCCG

(2) INFORMATION FOR SEQ ID NO:46:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 216 amino acids
 - (B) TYPE: amino acid (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

Met Gly Ser Ser His His His His His Ser Ser Gly Leu Val Pro $\frac{1}{5}$ 10 15

Arg Gly Ser His Met Gly Ser Val Val Ile Val Gly Arg Ile Ile Leu 20 25 30

Ser Pro Ala Gly Ile Thr Ala Tyr Ser Gln Gln Thr Arg Gly Leu Leu $35 \hspace{1cm} 40 \hspace{1cm} 45$

Gly Cys Ile Ile Thr Ser Leu Thr Gly Arg Asp Lys Asn Gln Val Glu 50 60

Gly Glu Val Gln Val Val Ser Thr Ala Thr Gln Ser Phe Leu Ala Thr $65 \\ 70 \\ 75 \\ 80$

Cys Val Asn Gly Val Cys Trp Thr Val Tyr His Gly Ala Gly Ser Lys $85 \hspace{1cm} 90 \hspace{1cm} 95$

Thr Leu Ala Gly Pro Lys Gly Pro Ile Thr Gln Met Tyr Thr Asn Val $100 \\ 105 \\ 110$

Asp Gln Asp Leu Val Gly Trp Gln Ala Pro Pro Gly Ala Arg Ser Leu 115 120 125

Thr Pro Cys Thr Cys Gly Ser Ser Asp Leu Tyr Leu Val Thr Arg His 130 135 140

Ala Asp Val Ile Pro Val Arg Arg Arg Gly Asp Ser Arg Gly Ser Leu 145 150 155 160

Leu Ser Pro Arg Pro Val Ser Tyr Leu Lys Gly Ser Ser Gly Gly Pro $165 \\ 170 \\ 170$

Leu Leu Cys Pro Ser Gly His Ala Val Gly Ile Phe Arg Ala Ala Val 180 185 190

Cys Thr Arg Gly Val Ala Lys Ala Val Asp Phe Val Pro Val Glu Ser 195 200 Met Glu Thr Thr Met Arg Ser * (2) INFORMATION FOR SEQ ID NO:47: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 3 amino acids (B) TYPE: amino acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear (ii) MOLECULE TYPE: peptide (xi) SEQUENCE DESCRIPTION: SEQ ID NO:47: Pro Ala Gly 1 (2) INFORMATION FOR SEQ ID NO:48: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 75 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear (ii) MOLECULE TYPE: cDNA (xi) SEQUENCE DESCRIPTION: SEQ ID NO:48: GATATACATA TGGGTTCTGT TGTTATTGTT GGTAGAATTA TTTTATCTCC TGCTGGTATC 60 ACGGCCTACT CCCAA 75 (2) INFORMATION FOR SEQ ID NO:49: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 36 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear (ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

- (2) INFORMATION FOR SEO ID NO:50:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 213 amino acids(B) TYPE: amino acid
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: protein
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:

Met Gly Ser Ser His His His His His Ser Ser Gly Leu Val Pro

1 5 10 15

Arg Gly Ser His Met Gly Ser Val Val Ile Val Gly Arg Ile Ile Leu $20 \\ 25 \\ 30$

Ser Pro Ala Gly Ile Thr Ala Tyr Ser Gln Gln Thr Arg Gly Leu Leu $35 \hspace{1.5cm} 40 \hspace{1.5cm} 45$

Gly Glu Val Gln Val Val Ser Thr Ala Thr Gln Ser Phe Leu Ala Thr 65 70 75 80

Cys Val Asn Gly Val Cys Trp Thr Val Tyr His Gly Ala Gly Ser Lys 85 90 95

Thr Leu Ala Gly Pro Lys Gly Pro Ile Thr Gln Met Tyr Thr Asn Val $100 \\ 105 \\ 110$

Asp Gln Asp Leu Val Gly Trp Gln Ala Pro Pro Gly Ala Arg Ser Leu 115 120 125

Thr Pro Cys Thr Cys Gly Ser Ser Asp Leu Tyr Leu Val Thr Arg His

Ala Asp Val Ile Pro Val Arg Arg Arg Gly Asp Ser Arg Gly Ser Leu 145 150 155 160

Leu Ser Pro Arg Pro Val Ser Tyr Leu Lys Gly Ser Ser Gly Gly Pro $165 \\ 170 \\ 175$

Cys Thr Arg Gly Val Ala Lys Ala Val Asp Phe Val Pro Val Glu Ser 195 200 205

Met Glu Thr Thr Met 210

(2) INFORMATION FOR SEQ ID NO:51:

(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 39 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear (ii) MOLECULE TYPE: cDNA (xi) SEQUENCE DESCRIPTION: SEO ID NO:51: CGGGGCCTAC TTGGTTGCAA GATCACTAGC CTTACAGGC 39 (2) INFORMATION FOR SEQ ID NO:52: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 39 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear (ii) MOLECULE TYPE: cDNA (xi) SEQUENCE DESCRIPTION: SEQ ID NO:52: GCCTGTAAGG CTAGTGATCT TGCAACCAAG TAGGCCCCG 39 (2) INFORMATION FOR SEQ ID NO:53: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 166 amino acids (B) TYPE: amino acid (D) TOPOLOGY: linear (ii) MOLECULE TYPE: protein (xi) SEQUENCE DESCRIPTION: SEQ ID NO:53: Met Ile Thr Ser Leu Thr Gly Arg Asp Lys Asn Gln Val Glu Gly Glu Val Gln Val Val Ser Thr Ala Thr Gln Ser Phe Leu Ala Thr Cys Val 20 Asn Gly Val Cys Trp Thr Val Tyr His Gly Ala Gly Ser Lys Thr Leu

35 40 45
Ala Gly Pro Lys Gly Pro Ile Thr Gln Met Tyr Thr Asn Val Asp Gln

60

55

50

Asp Leu Val Gly Trp Gln Ala Pro Pro Gly Ala Arg Ser Leu Thr Pro 65 70 75 80

Cys Thr Cys Gly Ser Ser Asp Leu Tyr Leu Val Thr Arg His Ala Asp 85 90 95

Pro Arg Pro Val Ser Tyr Leu Lys Gly Ser Ser Gly Gly Pro Leu Leu 115 120 125

Cys Pro Ser Gly His Ala Val Gly Ile Phe Arg Ala Ala Val Cys Thr 130 135 140

Arg Gly Val Ala Lys Ala Val Asp Phe Val Pro Val Glu Ser Met Glu 145 150 150 155 160

Thr Thr Met Arg Ser * 165

- (2) INFORMATION FOR SEQ ID NO:54:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 3 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: peptide
 - (xi) SEQUENCE DESCRIPTION: SEO ID NO:54:

Gly Gly Ser

- (2) INFORMATION FOR SEO ID NO:55:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 75 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: cDNA
 - (xi) SEQUENCE DESCRIPTION: SEO ID NO:55:

GATATACATA TGGGTTCTGT TGTTATTGTT GGTAGAATTA TTTTATCTGG TGGTTCTATC

- (2) INFORMATION FOR SEQ ID NO:56:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 36 base pairs (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: cDNA
 - (--) HODDCODD TILD. CDNA
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:

CTCAGCGAAT TCTCAAGACC GCATAGTAGT TTCCAT

36

- (2) INFORMATION FOR SEQ ID NO:57:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 216 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: protein
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:

Met Gly Ser Ser His His His His His Ser Ser Gly Leu Val Pro 1 5 10 15

Arg Gly Ser His Met Gly Ser Val Val Ile Val Gly Arg Ile Ile Leu $20 \\ 25 \\ 30$

Ser Gly Gly Ser Ile Thr Ala Tyr Ser Gln Gln Thr Arg Gly Leu Leu 35 40 45

Gly Glu Val Gln Val Val Ser Thr Ala Thr Gln Ser Phe Leu Ala Thr 65 70707580

Cys Val Asn Gly Val Cys Trp Thr Val Tyr His Gly Ala Gly Ser Lys $85 \\ 90 \\ 95$

Thr Leu Ala Gly Pro Lys Gly Pro Ile Thr Gln Met Tyr Thr Asn Val $100 \\ 0.05 \\ 105 \\ 110$

Asp Gln Asp Leu Val Gly Trp Gln Ala Pro Pro Gly Ala Arg Ser Leu 115 120 125

Thr Pro Cys Thr Cys Gly Ser Ser Asp Leu Tyr Leu Val Thr Arg His 130 135

Ala Asp Val Ile Pro Val Arg Arg Gly Asp Ser Arg Gly Ser Leu

155

160

150

- Met His Met His His His His His His Leu Val Pro Arg Gly Ser Ala

 1 5 10 15
- Pro Ile Thr Ala Tyr Ser Gln Gln Thr Arg Gly Leu Leu Gly Cys Lys 20 25 30
- Ile Thr Ser Leu Thr Gly Arg Asp Lys Asn Gln Val Glu Gly Glu Val \$35\$
- Gly Val Cys Trp Thr Val Tyr His Gly Ala Gly Ser Lys Thr Leu Ala 65 70 75 80
- Gly Pro Lys Gly Pro Ile Thr Gln Met Tyr Thr Asn Val Asp Gln Asp $85 \hspace{1.5cm} 90 \hspace{1.5cm} 95$
- Thr Cys Gly Ser Ser Asp Leu Tyr Leu Val Thr Arg His Ala Asp Val 115 120 125
- Ile Pro Val Arg Arg Arg Gly Asp Ser Arg Gly Ser Leu Leu Ser Pro 130 135 140
- Arg Pro Val Ser Tyr Leu Lys Gly Ser Ser Gly Gly Pro Leu Leu Cys 145 $$ $$ 150 $$ $$ 155
- Pro Ser Gly His Ala Val Gly Ile Phe Arg Ala Ala Val Cys Thr Arg 165 170 175
- Gly Val Ala Lys Ala Val Asp Phe Val Pro Val Glu Ser Met Glu Thr
- Thr Met Arg Ser Pro Val Phe Thr Asp Asn Ser Ser Pro Pro Ala Val 195 200
- Pro Gln Ser Phe Gln Val Ala His Leu His Ala Pro Thr Gly Ser Gly 210 215 220
- Lys Ser Thr Lys Val Pro Ala Ala Tyr Ala Ala Gln Gly Tyr Lys Val 225 230 235 240
- Leu Val Leu Asn Pro Ser Val Ala Ala Thr Leu Gly Phe Gly Ala Tyr \$245\$
- Met Ser Lys Ala His Gly Ile Asp Pro Asn Ile Arg Thr Gly Val Arg $260 \hspace{1.5cm} 265 \hspace{1.5cm} 265 \hspace{1.5cm} 270 \hspace{1.5cm}$
- Thr Ile Thr Thr Gly Ala Pro Val Thr Tyr Ser Thr Tyr Gly Lys Phe 275 280 285
- Leu Ala Asp Gly Gly Cys Ser Gly Gly Ala Tyr Asp Ile Ile Ile Cys 290 295 300
- Asp Glu Cys His Ser Thr Asp Ser Thr Thr Ile Leu Gly Ile Gly Thr

305					310					315					320
Val	Leu	Asp	Gln	Ala 325	Glu	Thr	Ala	Gly	Ala 330		Leu	Val	Val	Leu 335	
Thr	Ala	Thr	Pro 340	Pro	Gly	Ser	Val	Thr 345	Val	Pro	His	Pro	Asn 350		Glu
Glu	Val	Ala 355	Leu	Ser	Asn	Thr	Gly 360		Ile	Pro	Phe	Tyr 365		Lys	Ala
Ile	Pro 370	Ile	Glu	Ala	Ile	Arg 375	Gly	Gly	Arg	His	Leu 380		Phe	Cys	His
Ser 385	Lys	Lys	Lys	Cys	Asp 390	Glu	Leu	Ala	Ala	Lys 395	Leu	Ser	Gly	Leu	Gly 400
		Ala		405					410					415	
Thr	Ile	Gly	Asp 420	Val	Val	Val	Val	Ala 425	Thr	Asp	Ala	Leu	Met 430	Thr	Gly
Tyr	Thr	Gly 435	Asp	Phe	Asp	Ser	Val 440	Ile	Asp	Cys	Asn	Thr 445	Cys	Val	Thr
	450	Val				455					460				
Thr 465	Val	Pro	Gln	Asp	Ala 470	Val	Ser	Arg	Ser	Gln 475	Arg	Arg	Gly	Arg	Thr 480
		Gly		485					490					495	
		Gly	500					505					510		
		Ala 515					520					525			
	530	Tyr				535					540				
545		Trp			550					555					560
		Ser		565					570					575	
		Gln	580					585					590		
Trp		595					600					605			
Gly	Pro 610	Thr	Pro	Leu	Leu	Tyr 615	Arg	Leu	Gly	Ala	Val 620	Gln	Asn	Glu	Val

Thr Leu Thr His Pro Ile Thr Lys Tyr Ile Met Ala Cys Met Ser Ala 625 630 635 Asp Leu Glu Val Val Thr * Glu Phe Glu Leu Arg Arg Gln Ala Cys 645 650 Gly Arg Thr Arg Ala Pro Pro Pro Pro Pro Leu Arg 660 665 (2) INFORMATION FOR SEQ ID NO:61: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 39 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear (ii) MOLECULE TYPE: cDNA (xi) SEQUENCE DESCRIPTION: SEO ID NO:61: CGGGGCCTAC TTGGTTGCAA GATCACTAGC CTTACAGGC 39 (2) INFORMATION FOR SEQ ID NO:62: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 39 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear (ii) MOLECULE TYPE: cDNA (xi) SEQUENCE DESCRIPTION: SEQ ID NO:62: GCCTGTAAGG CTAGTGATCT TGCAACCAAG TAGGCCCCG 39 (2) INFORMATION FOR SEO ID NO:63: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 668 amino acids (B) TYPE: amino acid (D) TOPOLOGY: linear (ii) MOLECULE TYPE: protein (xi) SEQUENCE DESCRIPTION: SEQ ID NO:63:

Met His Met His His His His His Leu Val Pro Arg Gly Ser Ala

1	-			5					10)				15	i
Pro	Ile	Thr	Ala 20	Tyr	Ser	Gln	Gln	Thr 25		r Gly	Leu	Leu	Gly 30		Ile
Lys	Thr	Ser 35	Leu	Thr	Gly	Arg	Asp 40		Asn	Glr	. Val	Glu 45		Glu	Val
Gln	Val 50	Val	Ser	Thr	Ala	Thr 55		Ser	Phe	Leu	Ala 60		Cys	Val	Asn
Gly 65	Val	Cys	Trp	Thr	Val 70	Tyr	His	Gly	Ala	Gly 75	Ser	Lys	Thr	Leu	Ala 80
Gly	Pro	Lys	Gly	Pro 85	Ile	Thr	Gln	Met	Tyr 90	Thr	Asn	Val	Asp	Gln 95	Asp
Leu	Val	Gly	Trp 100	Gln	Ala	Pro	Pro	Gly 105	Ala	Arg	Ser	Leu	Thr 110	Pro	Cys
Thr	Cys	Gly 115	Ser	Ser	Asp	Leu	Tyr 120	Leu	Val	Thr	Arg	His 125	Ala	Asp	Val
Ile	Pro 130	Val	Arg	Arg	Arg	Gly 135	Asp	Ser	Arg	Gly	Ser 140	Leu	Leu	Ser	Pro
Arg 145	Pro	Val	Ser	Tyr	Leu 150	Lys	Gly	Ser	Ser	Gly 155	Gly	Pro	Leu	Leu	Cys 160
Pro	Ser	Gly	His	Ala 165	Val	Gly	Ile	Phe	Arg 170	Ala	Ala	Val	Cys	Thr 175	Arg
Gly	Val	Ala	Lys 180	Ala	Val	Asp	Phe	Val 185	Pro	Val	Glu	Ser	Met 190	Glu	Thr
Thr	Met	Arg 195	Ser	Pro	Val	Phe	Thr 200	Asp	Asn	Ser	Ser	Pro 205	Pro	Ala	Val
Pro	Gln 210	Ser	Phe	Gln	Val	Ala 215	His	Leu	His	Ala	Pro 220	Thr	Gly	Ser	Gly
Lys 225	Ser	Thr	Lys	Val	Pro 230	Ala	Ala	Tyr	Ala	Ala 235	Gln	Gly	Tyr	Lys	Val 240
Leu	Val	Leu	Asn	Pro 245	Ser	Val	Ala	Ala	Thr 250	Leu	Gly	Phe	Gly	Ala 255	Tyr
Met	Ser	Lys	Ala 260	His	Gly	Ile	Asp	Pro 265	Asn	Ile	Arg	Thr	Gly 270	Val	Arg
Thr	Ile	Thr 275	Thr	Gly	Ala	Pro	Val 280	Thr	Tyr	Ser	Thr	Tyr 285	Gly	Lys	Phe
Leu	Ala 290	Asp	Gly	Gly	Cys	Ser 295	Gly	Gly	Ala	Tyr	Asp 300	Ile	Ile	Ile	Cys
Asp 305	Glu	Cys	His	Ser	Thr 310	Asp	Ser	Thr	Thr	Ile 315	Leu	Gly	Ile	Gly	Thr 320

Val Leu Asp Gln Ala Glu Thr Ala Gly Ala Arg Leu Val Val Leu Ala 325 330 335

Thr Ala Thr Pro Pro Gly Ser Val Thr Val Pro His Pro Asn Ile Glu 340 345 350

Glu Val Ala Leu Ser Asn Thr Gly Glu Ile Pro Phe Tyr Gly Lys Ala $355 \hspace{1.5cm} 360 \hspace{1.5cm} 365 \hspace{1.5cm}$

Ile Pro Ile Glu Ala Ile Arg Gly Gly Arg His Leu Ile Phe Cys His $370 \hspace{1.5cm} 375 \hspace{1.5cm} 380 \hspace{1.5cm}$

Ser Lys Lys Lys Cys Asp Glu Leu Ala Ala Lys Leu Ser Gly Leu Gly 385 390 400

Ile Asn Ala Val Ala Tyr Tyr Arg Gly Leu Asp Val Ser Val Ile Pro 405 410 415

Thr Ile Gly Asp Val Val Val Val Ala Thr Asp Ala Leu Met Thr Gly 420 425 430

Tyr Thr Gly Asp Phe Asp Ser Val Ile Asp Cys Asn Thr Cys Val Thr 435 445

Gln Thr Val Asp Phe Ser Leu Asp Pro Thr Phe Thr Ile Glu Thr Thr 450 455 460

Thr Val Pro Gln Asp Ala Val Ser Arg Ser Gln Arg Arg Gly Arg Thr 465 470 475 480

Gly Arg Gly Arg Arg Gly Ile Tyr Arg Phe Val Thr Pro Gly Glu Arg 485 490 495

Pro Ser Gly Met Phe Asp Ser Ser Val Leu Cys Glu Cys Tyr Asp Ala 500 505 510

Gly Cys Ala Trp Tyr Glu Leu Thr Pro Ala Glu Thr Ser Val Arg Leu 515 520 525

Arg Ala Tyr Leu Asn Thr Pro Gly Leu Pro Val Cys Gln Asp His Leu 530 535 540

Glu Phe Trp Glu Ser Val Phe Thr Gly Leu Thr His Ile Asp Ala His 545 550 560

Phe Leu Ser Gln Thr Lys Gln Ala Gly Asp Asn Phe Pro Tyr Leu Val565 570 570 575

Ala Tyr Gln Ala Thr Val Cys Ala Arg Ala Gln Ala Pro Pro Pro Ser 580 585 590

Trp Asp Gln Met Trp Lys Cys Leu Ile Arg Leu Lys Pro Thr Leu His $595 \hspace{1.5cm} 600 \hspace{1.5cm} 605 \hspace{1.5cm}$

Gly Pro Thr Pro Leu Leu Tyr Arg Leu Gly Ala Val Gln Asn Glu Val 610 620

Thr Leu Thr His Pro Ile Thr Lys Tyr Ile Met Ala Cys Met Ser Ala 625 630 Asp Leu Glu Val Val Thr * (2) INFORMATION FOR SEQ ID NO:64: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 39 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear (ii) MOLECULE TYPE: cDNA (xi) SEQUENCE DESCRIPTION: SEQ ID NO:64: CGGGGCCTAC TTGGTTGCAT CAAGACTAGC CTTACAGGC 39 (2) INFORMATION FOR SEQ ID NO:65: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 39 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear (ii) MOLECULE TYPE: cDNA (xi) SEQUENCE DESCRIPTION: SEQ ID NO:65: GCCTGTAAGG CTAGTCTTGA TGCAACCAAG TAGGCCCCG 39 (2) INFORMATION FOR SEQ ID NO:66: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 668 amino acids (B) TYPE: amino acid (D) TOPOLOGY: linear (ii) MOLECULE TYPE: protein (xi) SEQUENCE DESCRIPTION: SEQ ID NO:66:

Pro Ile Thr Ala Tyr Ser Gln Gln Thr Arg Gly Leu Leu Gly Cys Ile \$20\$

Met His Met His His His His His Leu Val Pro Arg Gly Ser Ala

- Ile Thr Ser Leu Thr Gly Arg Asp Lys Asn Gln Val Glu Gly Glu Val $35 \ \ 40 \ \ 45$
- Gln Val Val Ser Thr Ala Thr Gln Ser Phe Leu Ala Thr Cys Val Asn 50 60
- Gly Val Cys Trp Thr Val Tyr His Gly Ala Gly Ser Lys Thr Leu Ala 65 70 75 80
- Gly Pro Lys Gly Pro Ile Thr Gln Met Tyr Thr Asn Val Asp Gln Asp 85 90 95
- Leu Val Gly Trp Gln Ala Pro Pro Gly Ala Arg Ser Leu Thr Pro Cys 100 105
- Thr Cys Gly Ser Ser Asp Leu Tyr Leu Val Thr Arg His Ala Asp Val 115 120 125
- Ile Pro Val Arg Arg Arg Gly Asp Ser Arg Gly Ser Leu Leu Ser Pro $130 \\ 135 \\ 140 \\ 140 \\ $
- Arg Pro Val Ser Tyr Leu Lys Gly Ser Ala Gly Gly Pro Leu Leu Cys 145 150 155 160
- Pro Ser Gly His Ala Val Gly Ile Phe Arg Ala Ala Val Cys Thr Arg 165 170 175
- Gly Val Ala Lys Ala Val Asp Phe Val Pro Val Glu Ser Met Glu Thr 180 185 190
- Thr Met Arg Ser Pro Val Phe Thr Asp Asn Ser Ser Pro Pro Ala Val
- Lys Ser Thr Lys Val Pro Ala Ala Tyr Ala Ala Gln Gly Tyr Lys Val 225 230 235 240
- Leu Val Leu Asn Pro Ser Val Ala Ala Thr Leu Gly Phe Gly Ala Tyr \$245\$ \$250\$ \$255\$
- Met Ser Lys Ala His Gly Ile Asp Pro Asn Ile Arg Thr Gly Val Arg 260 265 270
- Thr Ile Thr Thr Gly Ala Pro Val Thr Tyr Ser Thr Tyr Gly Lys Phe 275 280 285
- Leu Ala Asp Gly Gly Cys Ser Gly Gly Ala Tyr Asp Ile Ile Ile Cys 290 295 300
- Asp Glu Cys His Ser Thr Asp Ser Thr Thr Tle Leu Gly Ile Gly Thr 305 310 315
- Val Leu Asp Gln Ala Glu Thr Ala Gly Ala Arg Leu Val Val Leu Ala 325 \$330\$

- Thr Ala Thr Pro Pro Gly Ser Val Thr Val Pro His Pro Asn Ile Glu 340 345 350
- Glu Val Ala Leu Ser Asn Thr Gly Glu Ile Pro Phe Tyr Gly Lys Ala $355 \hspace{1.5cm} 360 \hspace{1.5cm} 365 \hspace{1.5cm}$
- Ile Pro Ile Glu Ala Ile Arg Gly Gly Arg His Leu Ile Phe Cys His 370 375 380
- Ser Lys Lys Lys Cys Asp Glu Leu Ala Ala Lys Leu Ser Gly Leu Gly 385 \$390\$
- Ile Asn Ala Val Ala Tyr Tyr Arg Gly Leu Asp Val Ser Val Ile Pro \$405\$
- Tyr Thr Gly Asp Phe Asp Ser Val Ile Asp Cys Asn Thr Cys Val Thr 435 440 445
- Gln Thr Val Asp Phe Ser Leu Asp Pro Thr Phe Thr Ile Glu Thr Thr $450 \ \ 455 \ \ 460$

- Pro Ser Gly Met Phe Asp Ser Ser Val Leu Cys Glu Cys Tyr Asp Ala 500 505 510

- Glu Phe Trp Glu Ser Val Phe Thr Gly Leu Thr His Ile Asp Ala His 545 550 555 560
- Phe Leu Ser Gln Thr Lys Gln Ala Gly Asp Asn Phe Pro Tyr Leu Val
- Ala Tyr Gln Ala Thr Val Cys Ala Arg Ala Gln Ala Pro Pro Pro Ser 580 585 590
- Trp Asp Gln Met Trp Lys Cys Leu Ile Arg Leu Lys Pro Thr Leu His 595 600 605
- Gly Pro Thr Pro Leu Leu Tyr Arg Leu Gly Ala Val Gln Asn Glu Val 610 615 620
- Asp Leu Glu Val Val Thr *

(2)	INFORM	ATION FO	R SEQ ID N	0:67:								
	(i) s	(A) LENG (B) TYPE (C) STRA	CHARACTERI TH: 40 base : nucleic a NDEDNESS: 1 LOGY: line	e pair acid single	rs.							
	(ii) M	OLECULE	TYPE: cDNA									
	(xi) S	EQUENCE I	DESCRIPTION	: SEQ	ID :	NO:6	7:					
CTC	CTACTTG	AAGGGCTC	TG CTGGTGGT	CC AC	TGCT	CTGC						40
(2)	INFORM	TION FOR	R SEQ ID NO	:68:								
		(A) LENG (B) TYPE (C) STRAI	CHARACTERIS TH: 40 base : nucleic a NDEDNESS: s LOGY: linea	pair cid ingle								
	(ii) M	OLECULE 1	PYPE: cDNA									
	(xi) SE	QUENCE D	ESCRIPTION	: SEQ	ID N	10:68	:					
GCA	SAGCAGT	GGACCACC2	AG CAGAGCCC	TT CA	AGTA	GAG						40
(2)	INFORMA	TION FOR	SEQ ID NO	:69:								
	(i)	(A) LE (B) TY	CHARACTER NGTH: 668 : PE: amino : POLOGY: li	amino acid		ls						
	(ii)	MOLECULE	TYPE: pro	cein								
	(xi)	SEQUENCE	DESCRIPTION	N: SE	Q ID	NO:	69:					
Met 1	His Met	His His 5	His His Hi	s His	Leu 10	Val	Pro	Arg	Gly	Ser 15	Ala	
Pro	Ile Thr	Ala Tyr 20	Ser Gln Gl	n Thr 25	Arg	Gly	Leu	Leu	Gly 30	Cys	Ile	

Ile Thr Ser Leu Thr Gly Arg Asp Lys Asn Gln Val Glu Gly Glu Val \$35\$

- Gly Val Cys Trp Thr Val Tyr His Gly Ala Gly Ser Lys Thr Leu Ala 65 70 75 80
- Gly Pro Lys Gly Pro Ile Thr Gln Met Tyr Thr Asn Val Asp Gln Asp 85 90 95
- Leu Val Gly Trp Gln Ala Pro Pro Gly Ala Arg Ser Leu Thr Pro Cys
- Thr Cys Gly Ser Ser Asp Leu Tyr Leu Val Thr Arg His Ala Asp Val $115 \\ 120 \\ 125$
- Ile Pro Val Arg Arg Arg Gly Asp Ser Arg Gly Ser Leu Leu Ser Pro 130 135 140
- Arg Pro Val Ser Tyr Leu Lys Gly Ser Ser Gly Gly Pro Leu Leu Cys 145 150 155 160
- Pro Ser Gly His Ala Val Gly Ile Phe Arg Ala Ala Val Cys Thr Arg 165 170 175
- Gly Val Ala Lys Ala Val Asp Phe Val Pro Val Glu Ser Met Glu Thr
- Thr Met Arg Ser Pro Val Phe Thr Asp Asn Ser Ser Pro Pro Ala Val 195 200 205
- Pro Gln Ser Phe Gln Val Ala His Leu His Ala Pro Thr Gly Ser Gly 210 215 220
- Lys Ser Thr Lys Val Pro Ala Ala Tyr Ala Ala Gln Gly Tyr Lys Val 225 230 235 240
- Leu Val Leu Asn Pro Ser Val Ala Ala Thr Leu Gly Phe Gly Ala Tyr
 245 250 255
- Met Ser Lys Ala His Gly Ile Asp Pro Asn Ile Arg Thr Gly Val Arg 260 265 270
- Thr Ile Thr Thr Gly Ala Pro Val Thr Tyr Ser Thr Tyr Gly Lys Phe \$275\$ \$280\$ \$285
- Leu Ala Asp Gly Gly Cys Ser Gly Gly Ala Tyr Asp Ile Ile Cys 290 295 300
- Asp Glu Cys His Ser Thr Asp Ser Thr Thr Ile Leu Gly Ile Gly Thr 305 310 315
- Val Leu Asp Gln Ala Glu Thr Ala Gly Ala Arg Leu Val Val Leu Ala 325 330 335
- Thr Ala Thr Pro Pro Gly Ser Val Thr Val Pro His Pro Asn Ile Glu 340 345 350
- Glu Val Ala Leu Ser Asn Thr Gly Glu Ile Pro Phe Tyr Gly Lys Ala

355 360 365

Ile Pro Ile Glu Ala Ile Arg Gly Gly Arg His Leu Ile Phe Cys His 370 375 380

Ser Lys Lys Lys Cys Asp Glu Leu Ala Ala Lys Leu Ser Gly Leu Gly 385 390 395

% Ile Asn Ala Val Ala Tyr Tyr Arg Gly Leu Asp Val Ser Val Ile Pro $405 \ \ 410 \ \ \ 415$

Tyr Thr Gly Asp Phe Asp Ser Val Ile Asp Cys Asn Thr Cys Val Thr 435 440 445

Gln Thr Val Asp Phe Ser Leu Asp Pro Thr Phe Thr Ile Glu Thr Thr 450 455 460 .

Thr Val Pro Gln Asp Ala Val Ser Arg Ser Gln Arg Arg Gly Arg Thr 465 470 475 480

Pro Ser Gly Met Phe Asp Ser Ser Val Leu Cys Glu Cys Tyr Asp Ala 500 505 510

Gly Cys Ala Trp Tyr Glu Leu Thr Pro Ala Glu Thr Ser Val Arg Leu 515 525 525

Arg Ala Tyr Leu Asn Thr Pro Gly Leu Pro Val Cys Gln Asp His Leu 530 535 540

Glu Phe Trp Glu Ser Val Phe Thr Gly Leu Thr His Ile Asp Ala His 545 550555555

Phe Leu Ser Gln Thr Lys Gln Ala Gly Asp Asn Phe Pro Tyr Leu Val 565 570 575

Ala Tyr Gln Ala Thr Val Cys Ala Arg Ala Gln Ala Pro Pro Pro Ser 580 585 590

Trp Asp Gln Met Trp Lys Cys Leu Ile Arg Leu Lys Pro Thr Leu His 595 600 605

Gly Pro Thr Pro Leu Leu Tyr Arg Leu Gly Ala Val Gln Asn Glu Val 610 $$\rm 615$$

Thr Leu Thr His Pro Ile Thr Lys Tyr Ile Met Ala Cys Met Ser Ala 625 $$ 630 $$ 635 $$ 640

Gly Arg Thr Arg Ala Pro Pro Pro Pro Pro Leu Arg

(2) Thromaton on one of the	
(2) INFORMATION FOR SEQ ID NO:70: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 35 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(ii) MOLECULE TYPE: cDNA	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:70:	
GTCCGTCATA CCAACTTCCG GAGACGTCGT TGTCG	35
(2) INFORMATION FOR SEQ ID NO:71:	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 35 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(ii) MOLECULE TYPE: cDNA	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:71: CGACAACGAC GTCTCCGGAA GTTGGTATGA CGGAC	35
(2) INFORMATION FOR SEQ ID NO:72:	
(i) SEQUENCE CHARACTERISTICS: (A) LEMGTH: 669 amino acids (B) TYPE: amino acid (D) TOPOLOGY: linear	
(ii) MOLECULE TYPE: protein	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:72:	
Met His Met His His His His His Leu Val Pro Arg Gly Ser Ala $1 \ 5 \ 10$	
Pro Ile Thr Ala Tyr Ser Gln Gln Thr Arg Gly Leu Leu Gly Cys Ile $20 \\ 20 \\ 30$	
Ile Thr Ser Leu Thr Gly Arg Asp Lys Asn Gln Val Glu Gly Glu Val 35 40 45	

Gln Val Val Ser Thr Ala Thr Gln Ser Phe Leu Ala Thr Cys Val Asn

	50					55					60				
Gly 65	Val	Cys	Trp	Thr	Val 70	Tyr	His	Gly	Ala	Gly 75	Ser	Lys	Thr	Leu	Ala 80
Gly	Pro	Lys	Gly	Pro 85	Ile	Thr	Gln	Met	Tyr 90	Thr	Asn	Val	Asp	G1n 95	Asp
Leu	Val	Gly	Trp 100	Gln	Ala	Pro	Pro	Gly 105	Ala	Arg	Ser	Leu	Thr 110	Pro	Cys
Thr	Cys	Gly 115	Ser	Ser	Asp	Leu	Tyr 120	Leu	Val	Thr	Arg	His 125	Ala	Asp	Val
Ile	Pro 130	Val	Arg	Arg	Arg	Gly 135	Asp	Ser	Arg	Gly	Ser 140	Leu	Leu	Ser	Pro
Arg 145	Pro	Val	Ser	Tyr	Leu 150	Lys	Gly	Ser	Ser	Gly 155	Gly	Pro	Leu	Leu	Cys 160
Pro	Ser	Gly	His	Ala 165	Val	Gly	Ile	Phe	Arg 170	Ala	Ala	Val	Cys	Thr 175	Arg
Gly	Val	Ala	Lys 180	Ala	Val	Asp	Phe	Val 185	Pro	Val	Glu	Ser	Met 190	Glu	Thr
Thr	Met	Arg 195	Ser	Pro	Val	Phe	Thr 200	Asp	Asn	Ser	Ser	Pro 205	Pro	Ala	Val
Pro	Gln 210	Ser	Phe	Gln	Val	Ala 215	His	Leu	His	Ala	Pro 220	Thr	Gly	Ser	Gly
Lys 225	Ser	Thr	Lys	Val	Pro 230	Ala	Ala	Tyr	Ala	Ala 235	Gln	Gly	Tyr	Lys	Val 240
Leu	Val	Leu	Asn	Pro 245	Ser	Val	Ala	Ala	Thr 250	Leu	Gly	Phe	Gly	A1a 255	Tyr
Met	Ser	Lys	Ala 260	His	Gly	Ile	Asp	Pro 265	Asn	Ile	Arg	Thr	Gly 270	Val	Arg
Thr	Ile	Thr 275	Thr	Gly	Ala	Pro	Val 280	Thr	Tyr	Ser	Thr	Tyr 285	Gly	Lys	Phe
Leu	Ala 290	Asp	Gly	Gly	Cys	Ser 295	Gly	Gly	Ala	Tyr	qaA 008	Ile	Ile	Ile	Cys
Asp 305	Glu	Cys	His	Ser	Thr 310	Asp	Ser	Thr	Thr	Ile 315	Leu	Gly	Ile	Gly	Thr 320
Va1	Leu	Asp	Gln	Ala 325	Glu	Thr	Ala	Gly	Ala 330	Arg	Leu	Val	Val	Leu 335	Ala
Thr	Ala		Pro	Pro	Gly	Ser	Val	Thr	Val	Pro	His	Pro	Asn	Ile	Glu

Glu Val Ala Leu Ser Asn Thr Gly Glu Ile Pro Phe Tyr Gly Lys Ala 355 360 365

- Ile Pro Ile Glu Ala Ile Arg Gly Gly Arg His Leu Ile Phe Cys His $370 \hspace{1.5cm} 375 \hspace{1.5cm} 380$
- Ser Lys Lys Lys Cys Asp Glu Leu Ala Ala Lys Leu Ser Gly Leu Gly 385 390 395 400
- Ile Asn Ala Val Ala Tyr Tyr Arg Gly Leu Asp Val Ser Val Ile Pro \$405\$
- Thr Ile Gly Asp Val Val Val Val Ala Thr Asp Ala Leu Met Thr Gly $420 \hspace{1.5cm} 425 \hspace{1.5cm} 430 \hspace{1.5cm}$
- Gln Thr Val Asp Phe Ser Leu Asp Pro Thr Phe Thr Ile Glu Thr Thr $450 \ \ \, 455 \ \ \, 460 \ \ \,$
- Thr Val Pro Gln Asp Ala Val Ser Arg Ser Gln Arg Arg Gly Arg Thr 465 470 475 480
- Gly Arg Gly Arg Gly Ile Tyr Arg Phe Val Thr Pro Gly Glu Arg
 485 490 495
- Pro Ser Gly Met Phe Asp Ser Ser Val Leu Cys Glu Cys Tyr Asp Ala 500 505 510
- Gly Cys Ala Trp Tyr Glu Leu Thr Pro Ala Glu Thr Ser Val Arg Leu 515 520 525
- Arg Ala Tyr Leu Asn Thr Pro Gly Leu Pro Val Cys Gln Asp His Leu 530 535 540
- Glu Phe Trp Glu Ser Val Phe Thr Gly Leu Thr His Ile Asp Ala His 545 550 555 560
- Phe Leu Ser Gln Thr Lys Gln Ala Gly Asp Asn Phe Pro Tyr Leu Val 565 570 575
- Ala Tyr Gln Ala Thr Val Cys Ala Arg Ala Gln Ala Pro Pro Pro Ser $580 \hspace{1.5cm} 585 \hspace{1.5cm} 590$
- Trp Asp Gln Met Trp Lys Cys Leu Ile Arg Leu Lys Pro Thr Leu His 595 600 605
- Gly Pro Thr Pro Leu Leu Tyr Arg Leu Gly Ala Val Gln Asn Glu Val 610 615 620
- Thr Leu Thr His Pro Ile Thr Lys Tyr Ile Met Ala Cys Met Ser Ala 625 630 635
- Asp Leu Glu Val Val Thr *
- (2) INFORMATION FOR SEQ ID NO:73:
 - (i) SEQUENCE CHARACTERISTICS:

		(A) LENGTH: 33 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(ii)	MOLECULE TYPE: cDNA	
	(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:73:	
ACTA	AAGTG	SC CGGCTGCCTA CGCAGCCCAA GGG	33
(2)	INFOR	RMATION FOR SEQ ID NO:74:	
	(i)	SEQUENCE CHARACTERISTICS: (A) LENGTH: 33 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(ii)	MOLECULE TYPE: cDNA	
	(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:74:	
CCCT	TGGGC	T GCGTAGGCAG CCGGCACTTT AGT	33
(2)	INFOR	MATION FOR SEQ ID NO:75:	
	(i)	SEQUENCE CHARACTERISTICS: (A) LENGTH: 39 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(ii)	MOLECULE TYPE: cDNA	
	(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:75:	
CGGG	GCCTA	C TTGGTTGCAA GATCACTAGC CTTACAGGC	39
(2)	INFOR	MATION FOR SEQ ID NO:76:	
	(i)	SEQUENCE CHARACTERISTICS: (A) LENGTH: 39 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(ii)	MOLECULE TYPE: cDNA	

	(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:76:	
GCCI	GTAAC	GG CTAGTGATCT TGCAACCAAG TAGGCCCCG	39
(2)	INFO	RMATION FOR SEQ ID NO:77:	
	(i)	SEQUENCE CHARACTERISTICS: (A) LENGTH: 39 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(ii)	MOLECULE TYPE: cDNA	
	(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:77:	
CGGG	GCCTA	C TTGGTTGCAT CAAGACTAGC CTTACAGGC	39
(2)	INFO	RMATION FOR SEQ ID NO:78:	
	(i)	SEQUENCE CHARACTERISTICS: (A) LENGTH: 39 base pairs (B) TYPE: nucleic acid (C) STRANDEDMESS: single (D) TOPOLOGY: linear	
	(ii)	MOLECULE TYPE: cDNA	
	(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:78:	
GCCT	GTAAG	G CTAGTCTTGA TGCAACCAAG TAGGCCCCG	39
(2)	INFOR	RMATION FOR SEQ ID NO:79:	
	(i)	SEQUENCE CHARACTERISTICS: (A) LENGTH: 38 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(ii)	MOLECULE TYPE: cDNA	
	(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:79:	

CGGGGCCTAC TTGGTTGCAA GAAGACTAGC CTTACAGG

(2) INFORMATION FOR SEQ ID NO:80:	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 39 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(ii) MOLECULE TYPE: cDNA	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:80:	
GCCTGTAAGG CTAGTCTTCT TGCAACCAAG TAGGCCCCG	39
(2) INFORMATION FOR SEQ ID NO:81:	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 39 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(ii) MOLECULE TYPE: cDNA	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:81: CGGGGCCTAC TTGGTTGCAA GATCACTAGC CTTACAGGC (2) INFORMATION FOR SEQ ID NO:82: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 39 base pairs (B) TYPE: nucleic acid (C) STRANDEDMESS: single (D) TOPOLOGY: linear	39
(ii) MOLECULE TYPE: cDNA	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:82:	
GCCTGTAAGG CTAGTGATCT TGCAACCAAG TAGGCCCCG	39
(2) INFORMATION FOR SEQ ID NO:83:	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 39 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:83:	
CGGGGCCTAC TTGGTTGCAT CAAGACTAGC CTTACAGGC	39
(2) INFORMATION FOR SEQ ID NO:84:	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 39 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(ii) MOLECULE TYPE: cDNA	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:84:	
GCCTGTAAGG CTAGTCTTGA TGCAACCAAG TAGGCCCCG	39
(2) INFORMATION FOR SEQ ID NO:85:	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 39 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(ii) MOLECULE TYPE: cDNA	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:85:	
CGGGGCCTAC TTGGTTGCAA GAAGACTAGC CTTACAGGC	39
(2) INFORMATION FOR SEQ ID NO:86:	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 39 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	

(ii) MOLECULE TYPE: cDNA

GCCTGTAAGG CTAGTCTTCT TGCAACCAAG TAGGCCCCG	39
(2) INFORMATION FOR SEQ ID NO:87:	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 96 base pairs (B) TYPE: nucleic acid (C) STRANDENNESS: single (D) TOPOLOGY: linear	
(ii) MOLECULE TYPE: cDNA	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:87:	
GATATACATA TGGCTTACTC TCTGACTACG GGTTCTGTTG TTATTGTTGG TAGAATTATT	60
TTATCTGGTA GTGGTAGTAT CACGGCCTAC TCCCAA	96
(2) INFORMATION FOR SEQ ID NO:88:	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 51 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(ii) MOLECULE TYPE: cDNA	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:88:	
GTGGTGGTGC TCGAGGCTGC CGCGCGCAC CAGCGTAACG ACCTCCAGGT C	51
(2) INFORMATION FOR SEQ ID NO:89:	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 96 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(ii) MOLECULE TYPE: cDNA	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:89:	
GATATACATA TGGCTTACTC TCTGACTACG GGTTCTGTTG TTATTGTTGG TAGAATTATT	60
TTATCTGGTA GTGGTAGTAT CACGGCCTAC TCCCAA	96
(2) INFORMATION FOR CEO ID NO.00.	

- (i) SEOUENCE CHARACTERISTICS:
 - (A) LENGTH: 50 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:90:

TGGTGGTGCT CGAGGCTGCC GCGCGGCACC AGCGTAACGA CCTCCAGGTC

(2) INFORMATION FOR SEQ ID NO:91:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 17 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: peptide
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:91:

Asp Thr Glu Asp Val Val Ala Cys Ser Met Ser Tyr Thr Trp Tyr Gly

Lys

- (2) INFORMATION FOR SEQ ID NO:92:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 651 base pairs
 - (B) TYPE: nucleic acid (D) TOPOLOGY: linear
 - (C) STRANDEDNESS: single
 - (ii) MOLECULE TYPE: cDNA
 - (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION: 1..651
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:92:

ATG GGC AGC AGC CAT CAT CAT CAT CAT CAC AGC AGC GGC CTG GTG CCG Met Gly Ser Ser His His His His His Ser Ser Gly Leu Val Pro

1			5			10			15	
	GGC Gly									96
	GGT Gly									144
	GGT Gly 50									192
	GGA Gly									240
	TGC Cys									288
	ACC Thr									336
	GAC Asp									384
	ACA Thr 130									432
	GCT Ala									480
	CTC Leu									528
	CTG Leu									576
	TGC Cys							Val 205		624
	ATG Met 210				Ser					651

(2) INFORMATION FOR SEQ ID NO:93:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 651 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
 - (ix) FEATURE:
 - (A) NAME/KEY: CDS

165

- (B) LOCATION: 1..651
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:93:

(XI) BEGOENCE DESCRIPTION. SEQ ID NO.33.															
		GGC Gly													48
		GGC Gly													96
		GGT Gly													144
		GGT Gly 50													192
		GGA Gly													240
		TGC Cys													288
		ACC Thr													336
		GAC Asp													384
		ACA Thr 130											Leu		432
		GCT Ala													480
		CTC Leu													528

170

CCA																
	CTG Leu															576
	TGC Cys														GAG Glu	624
	ATG Met 210						Ser									651
(2)	INFO	ORMAT	ON	FOR	SEQ	ID N	0:94	:								
	(i	(QUEN A) L B) T C) S D) T	ENGT YPE: TRAN	H: 6 nuc DEDN	51 b leic ESS:	ase aci sin	pair d	s							
	(ii) MO	LECU	LE T	YPE:	DNA	(ge	nomi	c)							
	(ix		ATUF A) N B) L	AME/												
	(xi) SE	QUEN	CE D	ESCR	IPTI	: NC	SEQ	ID N	0:94	:					
	GGC Gly													Val		48
	GGC								10					15		
Arg	Gly	AGC Ser							ATT					ATT		96
TCT		Ser AGT	His 20 GGT	Met AGT	Gly ATC	Ser ACG	Val GCC	Val 25 TAC	ATT Ile	Val CAA	Gly CAG	Arg ACG	Ile 30 CGG	ATT Ile GGC	Leu CTA	96 144
TCT Ser	Gly	AGT Ser 35	His 20 GGT Gly ATC	Met AGT Ser	Gly ATC Ile ACT	Ser ACG Thr	GCC Ala 40	Val 25 TAC Tyr ACA	ATT Ile TCC Ser	Val CAA Gln CGG	Gly CAG Gln GAC	Arg ACG Thr 45 AAG	Ile 30 CGG Arg	ATT Ile GGC Gly	Leu CTA Leu GTC	
TCT Ser CTT Leu	Gly GGT Gly GGT Gly	AGT Ser 35 TGC Cys	His 20 GGT Gly ATC Ile	Met AGT Ser AAG Lys	ATC Ile ACT Thr	ACG Thr AGC Ser 55	GCC Ala 40 CTT Leu	Val 25 TAC Tyr ACA Thr	ATT Ile TCC Ser GGC Gly	CAA Gln CGG Arg	CAG Gln GAC Asp 60	Arg ACG Thr 45 AAG Lys	Ile 30 CGG Arg AAC Asn	ATT Ile GGC Gly CAG Gln	CTA Leu GTC Val	144
TCT Ser CTT Leu GAG Glu 65	GGT Gly GGT Gly 50 GGA	AGT Ser 35 TGC Cys GAG Glu	His 20 GGT Gly ATC Ile GTT Val	Met AGT Ser AAG Lys CAG Gln GGC	ATC Ile ACT Thr GTG Val 70 GTG	ACG Thr AGC Ser 55 GTT Val	GCC Ala 40 CTT Leu TCC Ser	Val 25 TAC Tyr ACA Thr ACC	ATT Ile TCC Ser GGC Gly GCA Ala	CAA Gln CGG Arg ACA Thr 75	CAG Gln GAC Asp 60 CAA Gln	Arg ACG Thr 45 AAG Lys TCC Ser	CGG Arg AAC Asn TTC Phe	ATT Ile GGC Gly CAG Gln CTG Leu	CTA Leu GTC Val GCG Ala 80 TCA	144

	GAC Asp													384
	ACA Thr 130													432
	GCT Ala													480
	CTC Leu												•	528
	CTG Leu													576
	TGC Cys													624
	ATG Met 210						Ser							651
(2)	INFO) SE (QUEN A) L B) T	CE C ENGT YPE:	HARA H: 6	CTER 51 b	ISTI ase aci	CS: pair d	s					
		(C) S	TKAN	DEDN	ESS:	sin	gre						

- (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 1..651

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:95:

ATG GGC AGC AGC CAT CAT CAT CAT CAT CAC AGC AGC GGC CTG GTG CCG 48 Met Gly Ser Ser His His His His His Ser Ser Gly Leu Val Pro 1 10 CGC GGC AGC CAT ATG GGT TCT GTT GTT ATT GTT GGT AGA ATT ATT TTA 96 Arg Gly Ser His Met Gly Ser Val Val Ile Val Gly Arg Ile Ile Leu 20 25 30 TOT GGT AGT GGT AGT ATC ACG GCC TAC TCC CAA CAG ACG CGG GGC CTA 144

Ser	Gly	Ser 35	Gly	Ser	Ile	Thr	Ala 40	Tyr	Ser	Gln	Gln	Thr 45	Arg	Gly	Leu	
	GGT Gly 50															192
	GGA Gly															240
	TGC Cys														TCA Ser	288
	ACC Thr															336
	GAC Asp															384
	ACA Thr 130															432
	GCT Ala															480
	CTC Leu											_				528
	CTG Leu															576
	TGC Cys															624
	ATG Met 210															651

(2) INFORMATION FOR SEQ ID NO:96:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 650 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 1..650

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:96:

					л по с.		LOIV.	SEQ	י עד	WO: 91	0:					
ATG Met 1	GGC Gly	AGC Ser	AGC Ser	CAT His	His	CAT His	CAT His	CAT His	CAC His	Ser	AGC Sei	GGC Gly	CTG Let	GTG Val	CCG L Pro	48
CGC Arg	GGC Gly	AGC Ser	CAT His 20	Met	GGT Gly	TCT Ser	GTT Val	GTT Val	. Ile	GTT Val	GGT Gly	AGA Arg	ATT Ile	: Ile	TTA Leu	. 96
TCT Ser	GGT Gly	AGT Ser 35	GGT Gly	AGT Ser	ATC	ACG Thr	GCC Ala 40	Tyr	TCC Ser	CAA Gln	CAG Gln	ACG Thr 45	Arg	GGC Gly	CTA Leu	144
CTT Leu	GGT Gly 50	TGC Cys	ATC Ile	ATC Ile	ACT Thr	AGC Ser 55	CTT	ACA Thr	GGC Gly	CGG Arg	GAC Asp	AAG Lys	AAC Asn	CAG Gln	GTC Val	192
GAG Glu 65	GGA Gly	GAG Glu	GTT Val	CAG Gln	GTG Val 70	GTT Val	TCC	ACC Thr	GCA Ala	ACA Thr 75	CAA Gln	TCC Ser	TTC Phe	CTG Leu	GCG Ala 80	240
ACC Thr	TGC Cys	GTC Val	AAC Asn	GGC Gly 85	GTG Val	TGT Cys	TGG Trp	ACC Thr	GTT Val 90	TAC Tyr	CAT His	GGT Gly	GCT Ala	GGC Gly 95	TCA Ser	288
AAG Lys	ACC Thr	TTA Leu	GCC Ala 100	GGC Gly	CCA Pro	AAG Lys	GGG Gly	CCA Pro 105	ATC Ile	ACC Thr	CAG Gln	ATG Met	TAC Tyr 110	ACT Thr	AAT Asn	336
GTG Val	Asp	CAG Gln 115	GAC Asp	CTC Leu	GTC Val	GGC Gly	TGG Trp 120	CAG Gln	GCG Ala	CCC Pro	CCC Pro	GGG Gly 125	GCG Ala	CGT Arg	TCC Ser	384
TTG . Leu	ACA Thr 130	CCA Pro	TGC Cys	ACC Thr	TGT Cys	GGC Gly 135	AGC Ser	TCA Ser	GAC Asp	CTT Leu	TAC Tyr 140	TTG Leu	GTC Val	ACG Thr	AGA Arg	432
CAT 6 His . 145	GCT Ala	GAC Asp	gTC Val	Ile	CCG Pro 150	GTG Val	CGC Arg	CGG Arg	CGG Arg	GGC Gly 155	GAC Asp	AGT Ser	AGG Arg	GGG Gly	AGC Ser 160	480
CTG (Leu l	CTC Leu	TCC Ser	CCC Pro	AGG Arg 165	CCT Pro	GTC Val	TCC Ser	TAC Tyr	TTG Leu 170	AAG Lys	GGC Gly	TCT Ser	GCT Ala	GGT Gly 175	GGT Gly	528
CCA (CTG (Leu :	Leu	TGC Cys 180	CCT Pro	TCG Ser	GGG Gly	CAC His	GCT Ala 185	GTG Val	GGC Gly	ATC Ile	TTC Phe	CGG Arg 190	GCT Ala	GCC Ala	576
GTA 7	rgc :	ACC Thr	CGG Arg	GGG (GTT Val	GCG Ala	AAG Lys	GCG Ala	GTG Val	GAC Asp	TTT Phe	GTG Val	CCC Pro	GTA Val	GAG Glu	624

	ATG Met 210	Glu					Ser							650
(2)	INF	ORMA:	rion	FOR	SEQ	ID N	ю:97	:						
	(i	(A) L B) T C) S	ENGT YPE: TRAN	HARA H: 6 nuc IDEDN	50 b leic ESS:	ase aci sin	pair d	's					
	(ii) MC	LECU	LE T	YPE:	CDN	IA							
	(ix	(IAME/	KEY:									
	(xi) SE	QUEN	CE D	ESCR	IPTI	ON:	SEQ	ID N	0:97	:			
	GGC Gly													48
	GGC Gly													96
	GGT Gly													144
	GGT Gly 50													192
	GGA Gly													240
	TGC Cys													288
	ACC Thr													336
	GAC Asp													384

TTG ACA CCA TGC ACC TGT GGC AGC TCA GAC CTT TAC TTG GTC ACG AGA

rec	130	PIO	cys	Thr	Cys	135		Ser	Asp	Leu	140		. Va.	l Thi	r Arg	
CAT His 145	Ala	GAC Asp	GTC Val	ATT Ile	CCG Pro 150	GTG Val	CGC Arg	CGG Arg	CGG Arg	GGC G1y 155	Asp	AGT Ser	AGG	GGG Gly	AGC Ser 160	480
CTG Leu	CTC Leu	TCC Ser	CCC Pro	AGG Arg 165	CCT Pro	GTC Val	TCC Ser	TAC Tyr	TTG Leu 170	AAG Lys	GGC Gly	TCT	GCT Ala	GGT G13	GGT Gly	528
CCA Pro	CTG Leu	CTC Leu	TGC Cys 180	CCT Pro	TCG Ser	GGG Gly	CAC His	GCT Ala 185	Va1	GGC Gly	ATC Ile	TTC Phe	CGG Arg	Ala	GCC Ala	576
GTA Val	TGC Cys	ACC Thr 195	CGG Arg	GGG Gly	GTT Val	GCG Ala	AAG Lys 200	GCG Ala	GTG Val	GAC Asp	TTT Phe	GTG Val 205	Pro	GTA Val	GAG Glu	624
		GAA Glu					Ser									650
(2)	(ii (ix	() () () () MOI () FE ()	QUENCA) LI B) T' C) S' D) TC LECUI ATUR: A) NA B) LC	CE C. ENGTI YPE: FRANI OPOLI LE T' E: AME/I	HARA H: 6 nuc DEDN OGY: YPE: KEY: ION:	CTER 51 b leic ESS: lin cDN	ISTI ase aci sin ear A	CS: pair d gle								
a m.c		SEC														
Met 1	Gly	AGC Ser	Ser	His 5	His	His	His	His	His 10	AGC Ser	AGC Ser	GGC Gly	CTG Leu	GTG Val 15	CCG Pro	48
CGC Arg	GGC Gly	AGC Ser	CAT . His :	ATG Met	GGT Gly	TCT Ser	GTT Val	GTT Val 25	ATT Ile	GTT Val	GGT . Gly	AGA Arg	ATT Ile 30	ATT Ile	TTA Leu	96
TCT Ser	GGT Gly	AGT Ser 35	GGT A	AGT . Ser	ATC . Ile	ACG Thr	GCC Ala 40	TAC Tyr	TCC Ser	CAA Gln	CAG . Gln	ACG Thr 45	CGG Arg	GGC Gly	CTA Leu	144
CTT Leu	GGT Gly 50	TGC . Cys	ATC :	AAG I	ACT . Thr	AGC Ser 55	CTT . Leu	ACA Thr	GGC Gly	CGG Arg	GAC Asp	AAG Lys	AAC Asn	CAG Gln	GTC Val	192

GAG Glu 65	GGA Gly	GAG Glu	GTT Val	CAG Gln	GTG Val 70	Val	TCC	ACC Thr	GCA Ala	ACA Thr 75	CAA Gln	TCC Ser	TTC Phe	CTG Leu	GCG Ala 80	240
ACC Thr	TGC Cys	GTC Val	AAC Asn	GGC Gly 85	GTG Val	TGT Cys	TGG Trp	ACC Thr	GTT Val 90	Tyr	CAT	GGT Gly	GCT Ala	GGC Gly 95	TCA Ser	288
AAG Lys	ACC Thr	TTA Leu	GCC Ala 100	GGC Gly	CCA Pro	AAG Lys	GGG Gly	CCA Pro 105	Ile	ACC Thr	CAG Gln	ATG Met	TAC Tyr 110	Thr	AAT Asn	336
GTG Val	GAC Asp	CAG Gln 115	GAC Asp	CTC Leu	GTC Val	GGC Gly	TGG Trp 120	CAG Gln	GCG Ala	CCC Pro	CCC Pro	GGG Gly 125	GCG Ala	CGT Arg	TCC Ser	384
TTG Leu	ACA Thr 130	CCA Pro	TGC Cys	ACC Thr	TGT Cys	GGC Gly 135	AGC Ser	TCA Ser	GAC Asp	CTT Leu	TAC Tyr 140	Leu	GTC Val	ACG Thr	AGA Arg	432
	GCT Ala															480
	CTC Leu															528
CCA Pro	CTG Leu	CTC Leu	TGC Cys 180	CCT Pro	TCG Ser	GGG Gly	CAC His	GCT Ala 185	GTG Val	GGC Gly	ATC Ile	TTC Phe	CGG Arg 190	GCT Ala	GCC Ala	576
GTA Val	TGC Cys	ACC Thr 195	CGG Arg	GGG Gly	GTT Val	GCG Ala	AAG Lys 200	GCG Ala	GTG Val	GAC Asp	TTT Phe	GTG Val 205	CCC Pro	GTA Val	GAG Glu	624
	ATG Met 210							TGA *								651

(2) INFORMATION FOR SEQ ID NO:99:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 651 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..651

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:99:

GGC Gly							CCG Pro	48
GGC Gly								96
GGT Gly								144
GGT Gly 50								192
GGA Gly								240
TGC Cys								288
ACC Thr								336
GAC Asp								384
ACA Thr 130								432
GCT Ala								480
CTC Leu								528
CTG Leu								576
TGC Cys								624
ATG Met 210			Ser					651

(2) INFORMATION FOR SEQ ID NO:100:

(i) SEQUENCE	CHARACTERISTICS:
--------------------------------	------------------

- (A) LENGTH: 651 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 1..651

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:100:

ATG Met 1	Gly	AGC Ser	AGC Ser	CAT His	CAT	CAT	CAT	CAT His	CAC His	Ser	AGC Ser	GGC Gly	CTG Leu	GTG Val 15	CCG [.] Pro	48
CGC Arg	GGC Gly	AGC Ser	CAT His 20	ATG Met	GGT Gly	TCT Ser	GTT Val	GTT Val 25	ATT	GTT Val	GGT Gly	AGA Arg	ATT Ile 30	Ile	TTA Leu	96
TCT Ser	CCT Pro	GCT Ala 35	GGT Gly	GGT Gly	ATC Ile	ACG Thr	GCC Ala 40	TAC Tyr	TCC Ser	CAA Gln	CAG Gln	ACG Thr 45	CGG Arg	GGC Gly	CTA Leu	144
CTT Leu	GGT Gly 50	TGC Cys	ATC Ile	ATC Ile	ACT Thr	AGC Ser 55	CTT Leu	ACA Thr	GGC Gly	CGG Arg	GAC Asp 60	Lys	AAC Asn	CAG Gln	GTC Val	192
GAG Glu 65	GGA Gly	GAG Glu	GTT Val	CAG Gln	GTG Val 70	GTT Val	TCC Ser	ACC Thr	GCA Ala	ACA Thr 75	CAA Gln	TCC Ser	TTC Phe	CTG Leu	GCG Ala 80	240
ACC Thr	TGC Cys	GTC Val	AAC Asn	GGC Gly 85	GTG Val	TGT Cys	TGG Trp	ACC Thr	GTT Val 90	TAC Tyr	CAT His	GGT Gly	GCT Ala	GGC Gly 95	TCA Ser	288
AAG Lys	ACC Thr	TTA Leu	GCC Ala 100	GGC Gly	CCA Pro	AAG Lys	GGG Gly	CCA Pro 105	ATC Ile	ACC Thr	CAG Gln	ATG Met	TAC Tyr 110	ACT Thr	AAT Asn	336
GTG Val			GAC Asp													384
TTG Leu	ACA Thr 130	CCA Pro	TGC Cys	ACC Thr	TGT Cys	GGC Gly 135	AGC Ser	TCA Ser	GAC Asp	CTT Leu	TAC Tyr 140	TTG Leu	GTC Val	ACG Thr	AGA Arg	432
CAT His 145	GCT Ala	GAC Asp	GTC Val	ATT Ile	CCG Pro 150	GTG Val	CGC Arg	CGG Arg	CGG Arg	GGC Gly 155	GAC Asp	AGT Ser	AGG Arg	GGG Gly	AGC Ser 160	480

	CTG Leu	CTC	TCC Ser	CCC Pro	AGG Arg	Pro	GTC Val	TCC	TAC Tyr	Leu 170	ı Lys	GGC Gly	TCT	TCG Ser	GGT Gly 175	GGT Gly		528
	CCA Pro	CTG	Leu	TGC Cys 180	Pro	TCG Ser	GGG Gly	CAC	GCT Ala 185	. Val	GGC Gl	ATC	TTC Phe	CGG Arg	Ala	GCC Ala		576
	GTA Val	TGC Cys	ACC Thr 195	Arg	GGG Gly	GTT Val	GCG Ala	AAG Lys 200	Ala	GTG Val	GAC Asp	TTT Phe	GTG Val	Pro	GTA Val	GAG Glu		624
			Glu			ATG Met		Ser									,	651
	(2)	(ii) SE ((((() ()) MC	EQUENT (A) I (B) I (C) E	JCE (LENGT TYPE: STRAIN TOPOI LE T RE: HAME/	SEQ CHARA TH: 6 : nuc NDEDN LOGY: TYPE:	CTEF 51 h leic ESS: lir cDN	RIST] pase : aci sir near	CS: pair	:s						. •		
		(xi				TION:			SEQ	ID N	0:10	1:						
	ATG Met 1	GGC Gly	AGC Ser	AGC Ser	CAT His 5	CAT His	CAT His	CAT His	CAT His	CAC His 10	AGC Ser	AGC Ser	GGC Gly	CTG Leu	GTG Val 15	CCG Pro		48
	CGC Arg	GGC Gly	AGC Ser	CAT His 20	ATG Met	GGT Gly	TCT Ser	GTT Val	GTT Val 25	ATT Ile	GTT Val	GGT Gly	AGA Arg	ATT Ile 30	ATT Ile	TTA Leu		96
	PCT Ser	CCT Pro	GCT Ala 35	GGT Gly	GGT Gly	ATC Ile	ACG Thr	GCC Ala 40	TAC Tyr	TCC Ser	CAA Gln	CAG Gln	ACG Thr 45	CGG Arg	GGC Gly	CTA Leu		144
1	CTT Seu	GGT Gly 50	TGC Cys	AAG Lys	ATC Ile	ACT Thr	AGC Ser 55	CTT Leu	ACA Thr	GGC Gly	CGG Arg	GAC Asp 60	AAG Lys	AAC Asn	CAG Gln	GTC Val		192
0	GAG Glu 65	GGA Gly	GAG Glu	GTT Val	CAG Gln	GTG Val 70	GTT Val	TCC Ser	ACC Thr	GCA Ala	ACA Thr 75	CAA Gln	TCC Ser	TTC Phe	CTG Leu	GCG Ala 80		240
5	ACC	TGC Cys	GTC Val	AAC Asn	GGC Gly	GTG Val	TGT Cys	TGG Trp	ACC Thr	GTT Val	TAC Tyr	CAT His	GGT Gly	GCT Ala	GGC Gly	TCA Ser		288

AAG ACC TTA GCC GGC CCA AAG GGG CCA ATC ACC CAG ATG TAC ACT AAT

Lys Thr Leu Ala Gly Pro Lys Gly Pro Ile Thr Gln Met Tyr Thr Asn 100 105 110 GTG GAC CAG GAC CTC GTC GGC TGG CAG GCG CCC CCC GGG GCG CGT TCC

90

336

384

85

(ii) MOLECULE TYPE: cDNA

(A) NAME/KEY: CDS (B) LOCATION: 1..1995

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:102:

ATG GGC AGC AGC CAT CAT CAT CAT CAT CAC AGC AGC GGC CTG GTG CCG Met Gly Ser Ser His His His His His His Ser Ser Gly Leu Val Pro

10

15

(ix) FEATURE:

CG(GGC Gly	AGC Ser	CAT His	Met	GGT Gly	TCT Ser	GTT Val	GTT Val	. Ile	GTT Val	GGT Gly	AGA Arg	ATT Ile	Ile	TTA Leu		96
TC1 Ser	GGT Gly	AGT Ser 35	Gly	AGT	ATC	ACG Thr	GCC Ala	Tyr	TCC Ser	CAA Gln	CAG Gln	ACG Thr	Arg	GGC Gly	CTA Leu		144
CTI Leu	GGT Gly 50	Cys	ATC Ile	ATC	ACT Thr	AGC Ser 55	Leu	ACA Thr	GGC Gly	CGG Arg	GAC Asp	Lys	AAC Asn	CAG Gln	GTC Val		192
GAG Glu 65	GGA Gly	GAG Glu	GTT Val	CAG Gln	GTG Val 70	GTT Val	TCC Ser	ACC Thr	GCA Ala	ACA Thr 75	CAA Gln	TCC Ser	TTC Phe	CTG Leu	GCG Ala 80	*	240
ACC Thr	TGC Cys	GTC Val	AAC Asn	GGC Gly 85	GTG Val	TGT Cys	TGG Trp	ACC Thr	GTT Val 90	TAC Tyr	CAT His	GGT Gly	GCT Ala	GGC Gly 95	TCA Ser		288
AAG Lys	ACC Thr	TTA Leu	GCC Ala 100	GGC Gly	CCA Pro	AAG Lys	GGG Gly	CCA Pro 105	Ile	ACC Thr	CAG Gln	ATG Met	TAC Tyr 110	ACT Thr	AAT Asn		336
GTG Val	GAC Asp	CAG Gln 115	GAC Asp	CTC Leu	GTC Val	GGC Gly	TGG Trp 120	Gln	GCG Ala	CCC Pro	CCC Pro	GGG Gly 125	GCG Ala	CGT Arg	TCC Ser		384
TTG Leu	ACA Thr 130	CCA Pro	TGC Cys	ACC Thr	TGT Cys	GGC Gly 135	AGC Ser	TCA Ser	GAC Asp	CTT Leu	TAC Tyr 140	TTG Leu	GTC Val	ACG Thr	AGA Arg		432
CAT His 145	GCT Ala	GAC Asp	GTC Val	ATT Ile	CCG Pro 150	GTG Val	CGC Arg	CGG Arg	CGG Arg	GGC Gly 155	GAC Asp	AGT Ser	AGG Arg	GGG Gly	AGC Ser 160		480
CTG Leu	CTC Leu	TCC Ser	CCC Pro	AGG Arg 165	CCT Pro	GTC Val	TCC Ser	TAC Tyr	TTG Leu 170	AAG Lys	GGC Gly	TCT Ser	TCG Ser	GGT Gly 175	GGT Gly		528
CCA Pro	CTG Leu	CTC Leu	TGC Cys 180	CCT Pro	TCG Ser	GGG Gly	CAC His	GCT Ala 185	GTG Val	GGC Gly	ATC Ile	TTC Phe	CGG Arg 190	GCT Ala	GCC Ala		576
GTA Val	TGC Cys	ACC Thr 195	CGG Arg	GGG Gly	GTT Val	GCG Ala	AAG Lys 200	GCG Ala	GTG Val	GAC Asp	TTT Phe	GTG Val 205	CCC Pro	GTA Val	GAG Glu		624
TCC Ser	ATG Met 210	GAA Glu	ACT Thr	ACT Thr	ATG Met	CGG Arg 215	TCT Ser	CCG Pro	GTC Val	TTC Phe	ACG Thr 220	GAC Asp	AAC Asn	TCA Ser	TCC Ser		672
CCC Pro 225	CCG Pro	GCC Ala	GTA Val	CCG Pro	CAG Gln 230	TCA Ser	TTT Phe	CAA Gln	GTG Val	GCC Ala 235	CAC His	CTA Leu	CAC His	GCT Ala	CCC Pro 240		720
	GGC Gly																768

	245	250		255
Gly Tyr Lys V			GTT GCC GCT ACC Val Ala Ala Thr 270	Leu Gly
			ATT GAC CCC AAC Ile Asp Pro Asn 285	
			CCC GTC ACA TAC Pro Val Thr Tyr 300	
			TCT GGG GGC GCT Ser Gly Gly Ala 315	
			GAC TCG ACT ACA Asp Ser Thr Thr	
Gly Ile Gly T			ACG GCT GGA GCG Thr Ala Gly Ala 350	Arg Leu
			TCG GTC ACC GTG Ser Val Thr Val 365	
			ACT GGA GAG ATC Thr Gly Glu Ile 380	
			AGG GGG GGA AGG Arg Gly Gly Arg 395	
			GAG CTC GCC GCA Glu Leu Ala Ala	
Ser Gly Leu G			TAC CGG GGG CTC Tyr Arg Gly Leu 430	Asp Val
			GTC GTG GCA ACA Val Val Ala Thr 445	
			TCA GTG ATC GAC Ser Val Ile Asp 460	
			TTG GAT CCC ACC Leu Asp Pro Thr 475	

ATT Ile	GAG Glu	ACG Thr	ACG Thr	ACC Thr 485	Val	CCT Pro	CAA Gln	GAC Asp	GCA Ala 490	Val	TCG Ser	CGC Arg	TCG Ser	CAG Gln 495	Arg	1488
CGG Arg	GGT	AGG Arg	ACT Thr 500	Gly	AGG Arg	GGT	AGG Arg	AGA Arg 505	Gly	ATC	TAC Tyr	AGG Arg	TTT Phe 510	GTG Val	ACT Thr	1536
CCG Pro	GGA Gly	GAA Glu 515	Arg	CCC	TCG Ser	GGC	ATG Met 520	Phe	GAT Asp	TCC Ser	TCG Ser	GTC Val 525	CTG Leu	TGT Cys	GAG Glu	1584
TGC Cys	TAT Tyr 530	GAC Asp	GCG Ala	GGC Gly	TGT Cys	GCT Ala 535	TGG Trp	TAC Tyr	GAG Glu	CTC Leu	ACC Thr 540	Pro	GCC Ala	GAG Glu	ACC Thr	1632
TCG Ser 545	GTT Val	AGG Arg	TTG Leu	CGG Arg	GCC Ala 550	TAC Tyr	CTG Leu	AAC Asn	ACA Thr	CCA Pro 555	GGG Gly	TTG Leu	CCC Pro	GTT Val	TGC Cys 560	1680
CAG Gln	GAC Asp	CAC His	CTG Leu	GAG Glu 565	TTC Phe	TGG Trp	GAG Glu	AGT Ser	GTC Val 570	TTC Phe	ACA Thr	GGC Gly	CTC Leu	ACC Thr 575	CAT His	1728
ATA Ile	GAT Asp	GCA Ala	CAC His 580	TTC Phe	TTG Leu	TCC Ser	CAG Gln	ACC Thr 585	AAG Lys	CAG Gln	GCA Ala	GGA Gly	GAC Asp 590	AAC Asn	TTC Phe	1776
CCC Pro	TAC Tyr	CTG Leu 595	GTA Val	GCA Ala	TAC Tyr	CAA Gln	GCC Ala 600	ACG Thr	GTG Val	TGC Cys	GCC Ala	AGG Arg 605	GCT Ala	CAG Gln	GCC Ala	1824
CCA Pro	CCT Pro 610	CCA Pro	TCA Ser	TGG Trp	GAT Asp	CAA Gln 615	ATG Met	TGG Trp	AAG Lys	TGT Cys	CTC Leu 620	ATA Ile	CGG Arg	CTG Leu	AAA Lys	1872
CCT Pro 625	ACG Thr	CTG Leu	CAC His	GGG Gly	CCA Pro 630	ACA Thr	CCC Pro	TTG Leu	CTG Leu	TAC Tyr 635	AGG Arg	CTG Leu	GGA Gly	GCC Ala	GTC Val 640	1920
CAA Gln	AAT Asn	GAG Glu	GTC Val	ACC Thr 645	CTC Leu	ACC Thr	CAC His	CCC Pro	ATA Ile 650	ACC Thr	AAA Lys	TAC Tyr	ATC Ile	ATG Met 655	GCA	1968
		TCG Ser							ACT							1998

665

(2) INFORMATION FOR SEQ ID NO:103:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1998 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix)	FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..1997

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..1995

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:103:

ATG GGC AGC AGC CAT CAT CAT CAT CAT CAC AGC AGC GGC CTG GTG CCG Met Gly Ser Ser His His His His His His Ser Ser Gly Leu Val Pro 1 15 CGC GGC AGC CAT ATG GGT TCT GTT GTT ATT GTT GGT AGA ATT ATT TTA- Arg Gly Ser His Met Gly Ser Val Val Ile Val Gly Arg Ile Ile Leu 20 25 TCT GGT AGT GGT AGT ATC AGG GCC TAC TCC CAA CAG ACG CGG GGC CTA Ser Gly Ser Gly Ser Ile Thr Ala Tyr Ser Gln Gln Thr Arg Gly Leu 35 CTT GGT TGC AAG ATC ACT ACC CTT ACA GGC CGG GAC AAG AAC CAG GTC Leu Gly Cys Lys Ile Thr Ser Leu Thr Gly Arg Asp Lys Asn Gln Val 50 55 60	
Arg Gly Ser His Met Gly Ser Val Val Ile Val Gly Arg Ile Ile Leu 20 25 30 TCT GGT AGT GGT AGT ATC ACG GCC TAC TCC CAA CAG ACG CGG GGC CTA Ser Gly Ser Gly Ser Ile Thr Ala Tyr Ser Gln Gln Thr Arg Gly Leu 35 40 CTT GGT TGC AAG ATC ACT AGC CTT ACA GGC CGG GAC AAG AAC CAG GTC Leu Gly Cys Lys Ile Thr Ser Leu Thr Gly Arg Asp Lys Asn Gln Val	48
Ser Gly Ser Gly Ser Ile Thr Ala Tyr Ser Gln Gln Thr Arg Gly Leu 35 40 45 CTT GGT TGC AAG ATC ACT AGC CTT ACA GGC CGG GAC AAG AAC CAG GTC Leu Gly Cys Lys Ile Thr Ser Leu Thr Gly Arg Asp Lys Asn Gln Val	96
Leu Gly Cys Lys Ile Thr Ser Leu Thr Gly Arg Asp Lys Asn Gln Val	144
	192
GAG GGA GAG GTT CAG GTG GTT TCC ACC GCA ACA CAA TCC TTC CTG GCG Glu Gly Glu Val Gln Val Val Ser Thr Ala Thr Gln Ser Phe Leu Ala 65 70 80	240
ACC TGC GTC AAC GGC GTG TGT TGG ACC GTT TAC CAT GGT GCT GGC TCA Thr Cys Val Asn Gly Val Cys Trp Thr Val Tyr His Gly Ala Gly Ser 85 90 95	288
AAG ACC TTA GCC GGC CCA AAG GGG CCA ATC ACC CAG ATG TAC ACT AAT Lys Thr Leu Ala Gly Pro Lys Gly Pro Ile Thr Gln Met Tyr Thr Asn 100 105 110	336
GTG GAC CAG GAC CTC GTC GGC TGG CAG GCG CCC CCC GGG GCG CGT TCC Val Asp Gln Asp Leu Val Gly Trp Gln Ala Pro Pro Gly Ala Arg Ser 115 120 125	384
TTG ACA CCA TGC ACC TGT GGC AGC TCA GAC CTT TAC TTG GTC ACG AGA Leu Thr Pro Cys Thr Cys Gly Ser Ser Asp Leu Tyr Leu Val Thr Arg 130 135 140	432
CAT GCT GAC GTC ATT CCG GTG CGC CGG GGC GAC AGT AGG GGG AGC His Ala Asp Val Ile Pro Val Arg Arg Arg Gly Asp Ser Arg Gly Ser 145 150 155 160	480

CTG CTC TCC CCC AGG CCT GTC TCC TAC TTG AAG GGC TCT TCG GGT GGT

Leu Leu Ser Pro Arg Pro Val Ser Tyr Leu Lys Gly Ser Ser Gly Gly

170

165

528

CCA Pro	. CTG	CTC Leu	TGC Cys 180	Pro	TCG Ser	GGG Gly	CAC His	GCT Ala 185	Va]	GGC Gly	ATC	TTC Phe	CGG Arg	Ala	GCC Ala	576
GTA Val	TGC Cys	ACC Thr 195	Arg	GGG Gly	GTT Val	GCG Ala	AAG Lys 200	Ala	GTG Val	GAC Asp	TTT Phe	GTG Val 205	Pro	GTA Val	GAG Glu	624
TCC	ATG Met 210	Glu	ACT Thr	ACT Thr	ATG Met	CGG Arg 215	Ser	CCG Pro	GTC Val	TTC Phe	ACG Thr 220	Asp	AAC Asn	TCA Ser	TCC Ser	672
CCC Pro 225	CCG Pro	GCC Ala	GTA Val	CCG Pro	CAG Gln 230	Ser	TTT Phe	CAA Gln	GTG Val	GCC Ala 235	His	CTA Leu	CAC	GCT Ala	CCC Pro 240	720
ACT Thr	GGC Gly	AGC Ser	GGC Gly	AAG Lys 245	AGT Ser	ACT Thr	AAA Lys	GTG Val	CCG Pro 250	Ala	GCA Ala	TAT Tyr	GCA Ala	GCC Ala 255	Gln	768
GGG Gly	TAC Tyr	AAG Lys	GTG Val 260	CTC Leu	GTC Val	CTC Leu	AAT Asn	CCG Pro 265	TCC Ser	GTT Val	GCC Ala	GCT Ala	ACC Thr 270	TTA Leu	GGG Gly	816
TTT Phe	GGG Gly	GCG Ala 275	TAT Tyr	ATG Met	TCT Ser	AAG Lys	GCA Ala 280	CAC His	GGT Gly	ATT Ile	GAC Asp	CCC Pro 285	AAC Asn	ATC Ile	AGA Arg	864
ACT Thr	GGG Gly 290	GTA Val	AGG Arg	ACC Thr	ATT Ile	ACC Thr 295	ACA Thr	GGC Gly	GCC Ala	CCC Pro	GTC Val 300	ACA Thr	TAC Tyr	TCT Ser	ACC Thr	912
TAT Tyr 305	GGC Gly	AAG Lys	TTT Phe	CTT Leu	GCC Ala 310	GAT Asp	GGT Gly	GGT Gly	TGC Cys	TCT Ser 315	GGG Gly	GGC Gly	GCT Ala	TAT Tyr	GAC Asp 320	960
ATC Ile	ATA Ile	ATA Ile	TGT Cys	GAT Asp 325	GAG Glu	TGC Cys	CAT His	TCA Ser	ACT Thr 330	GAC Asp	TCG Ser	ACT Thr	ACA Thr	ATC Ile 335	TTG Leu	1008
GGC Gly	ATC Ile	GGC Gly	ACA Thr 340	GTC Val	CTG Leu	GAC Asp	CAA Gln	GCG Ala 345	GAG Glu	ACG Thr	GCT Ala	GGA Gly	GCG Ala 350	CGG Arg	CTT Leu	1056
GTC Val	GTG Val	CTC Leu 355	GCC Ala	ACC Thr	GCT Ala	ACG Thr	CCT Pro 360	CCG Pro	GGA Gly	TCG Ser	GTC Val	ACC Thr 365	GTG Val	CCA Pro	CAC His	1104
CCA Pro	AAC Asn 370	ATC Ile	GAG Glu	GAG Glu	GTG Val	GCC Ala 375	CTG Leu	TCT Ser	AAT Asn	ACT Thr	GGA Gly 380	GAG Glu	ATC Ile	CCC Pro	TTC Phe	1152
TAT Tyr 385	GGC Gly	AAA Lys	GCC Ala	ATC Ile	CCC Pro 390	ATT Ile	GAA Glu	GCC Ala	ATC Ile	AGG Arg 395	GGG Gly	GGA Gly	AGG Arg	CAT His	CTC Leu 400	1200
ATT Ile	TTC Phe	TGT Cys	CAT His	TCC Ser	AAG Lys	AAG Lys	AAG Lys	TGC Cys	GAC Asp	GAG Glu	CTC Leu	GCC Ala	GCA Ala	AAG Lys	CTG Leu	1248

405 410 415

				405	,				410)				41	5	
TCA Ser	GGC	CTC Lev	GGA Gly 420	r Il∈	AAC Asr	GCT Ala	GTG Val	GCG L Ala 425	a Tyr	TAC	CGG Arg	GGG GGG	CTC Leu 430	ı Ası	GTG Val	1296
TCC	GTC Val	11e 435	Pro	ACT Thr	ATC	GGA Gly	GAC Asp 440	Val	GTT Val	GTC Val	GTG Val	GCA Ala 445	Thr	GAC Asp	GCT Ala	1344
CTG Leu	ATG Met 450	Thr	GGC	TAT	ACG Thr	GGC Gly 455	Asp	TTT Phe	GAC Asp	TCA Ser	GTG Val		GAC Asp	TGT Cys	AAC Asn	1392
ACA Thr 465	TGT	GTC Val	ACC Thr	CAG Gln	ACA Thr 470	GTC Val	GAC Asp	TTC Phe	AGC Ser	TTG Leu 475	Asp	CCC Pro	ACC Thr	TTC Phe	ACC Thr 480	1440
ATT Ile	GAG Glu	ACG Thr	ACG Thr	ACC Thr 485	Val	CCT Pro	CAA Gln	GAC Asp	GCA Ala 490	Val	TCG Ser	CGC Arg	TCG Ser	CAG Gln 495		1488
CGG Arg	GGT Gly	AGG Arg	ACT Thr 500	GGC Gly	AGG Arg	GGT Gly	AGG Arg	AGA Arg 505	Gly	ATC Ile	TAC Tyr	AGG Arg	TTT Phe 510	GTG Val	ACT Thr	1536
Pro	Gly	Glu 515	Arg	Pro	Ser	Gly	Met 520	Phe	Asp	Ser	Ser	GTC Val 525	Leu	Cys	Glu	1584
TGC Cys	TAT Tyr 530	GAC Asp	GCG Ala	GGC Gly	TGT Cys	GCT Ala 535	TGG Trp	TAC Tyr	GAG Glu	CTC Leu	ACC Thr 540	CCC Pro	GCC Ala	GAG Glu	ACC Thr	1632
TCG Ser 545	GTT Val	AGG Arg	TTG Leu	CGG Arg	GCC Ala 550	TAC Tyr	CTG Leu	AAC Asn	ACA Thr	CCA Pro 555	GGG Gly	TTG Leu	CCC Pro	GTT Val	TGC Cys 560	1680
CAG G1n	GAC Asp	CAC His	CTG Leu	GAG Glu 565	TTC Phe	TGG Trp	GAG Glu	AGT Ser	GTC Val 570	TTC Phe	ACA Thr	GGC Gly	CTC Leu	ACC Thr 575	CAT His	1728
ATA Ile	GAT Asp	GCA Ala	CAC His 580	TTC Phe	TTG Leu	TCC Ser	CAG Gln	ACC Thr 585	AAG Lys	CAG Gln	GCA Ala	GGA Gly	GAC Asp 590	AAC Asn	TTC Phe	1776
CCC Pro	TAC Tyr	CTG Leu 595	GTA Val	GCA Ala	TAC Tyr	CAA Gln	GCC Ala 600	ACG Thr	GTG Val	TGC Cys	GCC Ala	AGG Arg 605	GCT Ala	CAG Gln	GCC Ala	1824
CCA Pro	CCT Pro 610	CCA Pro	TCA Ser	TGG Trp	GAT Asp	CAA Gln 615	ATG Met	TGG Trp	AAG Lys	TGT Cys	CTC Leu 620	ATA Ile	CGG Arg	CTG Leu	AAA Lys	1872
CCT Pro 625	ACG Thr	CTG Leu	CAC His	GGG Gly	CCA Pro 630	ACA Thr	CCC Pro	TTG Leu	CTG Leu	TAC Tyr 635	AGG Arg	CTG Leu	GGA Gly	GCC Ala	GTC Val 640	1920

CAA Glr	AAT ASI	GAC	G GTO	2 ACC 1 Thi 645	: Lei	C ACC	C CAC	C CCC	0 II:	e Thi	C AAA	A TAC	C ATO	ATC Me 65	G GCA t Ala 5	1968
				r GAC a Asl					1	,						1998
(2)	INF	ORMA	TION	FOR	SEQ	ID	NO:1	04:								
	(:		(A) (B) (C)	NCE LENG' TYPE STRAI	FH: : nu NDEDI	1998 clei NESS	base c ac: sin	e pa id	irs							
	(ii) M	OLEC	ULE '	PYPE	: cDi	NA									
	(i:			RE: NAME,				5							•	
	(xi) SE	QUEN	ICE E	ESCF	IPTI	ON:	SEQ	ID N	JO:10	4:					
ATG Met 1	GGC Gly	AGC Ser	AGC Ser	CAT His	CAT His	CAT His	CAT His	CAT	CAC His	Ser	AGC Ser	GGC Gly	CTG Leu	GTG Val	CCG Pro	48
CGC Arg	GGC Gly	AGC Ser	CAT His 20	ATG Met	GGT Gly	TCT Ser	GTT Val	GTT Val 25	ATT Ile	GTT Val	GGT Gly	AGA Arg	ATT Ile 30	Ile	TTA Leu	96
TCT Ser	GGT Gly	AGT Ser 35	GGT Gly	AGT Ser	ATC Ile	ACG Thr	GCC Ala 40	TAC Tyr	TCC Ser	CAA Gln	CAG Gln	ACG Thr 45	CGG Arg	GGC Gly	CTA Leu	144
CTT Leu	GGT Gly 50	TGC Cys	ATC Ile	AAG Lys	ACT Thr	AGC Ser 55	CTT Leu	ACA Thr	GGC Gly	CGG Arg	GAC Asp 60	Lys	AAC Asn	CAG Gln	GTC Val	192
GAG G1u 65	GGA Gly	GAG Glu	GTT Val	CAG Gln	GTG Val 70	GTT Val	TCC Ser	ACC Thr	GCA Ala	ACA Thr 75	CAA Gln	TCC Ser	TTC Phe	CTG Leu	GCG Ala 80	240
ACC Thr	TGC Cys	GTC Val	AAC Asn	GGC Gly 85	GTG Val	TGT Cys	TGG Trp	ACC Thr	GTT Val 90	TAC Tyr	CAT His	GGT Gly	GCT Ala	GGC Gly 95	Ser	288
AAG Lys	ACC Thr	TTA Leu	GCC Ala 100	GGC Gly	CCA Pro	AAG Lys	GGG Gly	CCA Pro 105	ATC Ile	ACC Thr	CAG Gln	ATG Met	TAC Tyr 110	ACT Thr	AAT Asn	336
GTG Val	GAC Asp	CAG Gln	GAC Asp	CTC Leu	GTC Val	GGC Gly	TGG	CAG Gln	GCG Ala	CCC	CCC	GGG G1v	GCG Ala	CGT	TCC	384

115 120 125

TTG	ACA	CCA	TGC	ACC	TGT	GGC	AGC	TCA	GAC	CTT	TAC	TTG	GTC	ACG	AGA	432
Leu	130	Pro	Cys	Thr	Cys	Gly 135	Ser	Ser	Asp	Leu	140	Leu)	Val	Thr	Arg	
CAT	GCT	GAC	GTC Val	ATT	CCG	GTG Val	CGC	CGG	CGG	GGC	GAC	AGT	AGG	GGG	AGC Ser	480
145					150					155					160	
CTG Leu	CTC Leu	TCC	CCC Pro	AGG Arg	CCT	GTC Val	TCC	TAC	TTG	AAG	GGC Glv	TCT	TCG	GGT Glv	GGT Gly	528
				165				-	170		2			175		
CCA	CTG	CTC	TGC	CCT	TCG	GGG	CAC	GCT	GTG	GGC	ATC	TTC	CGG	GCT	GCC Ala	576
110	Бец	пец	180	PIO	ser	GIĀ	HIS	185	vaı	GIY	ile	Phe	190		Ala	
GTA Val	TGC	ACC	CGG	GGG Glv	GTT Val	GCG Ala	AAG	GCG	GTG	GAC	TTT	GTG	CCC	GTA	GAG Glu	624
		195					200					205				
TCC	ATG Met	GAA Glu	ACT	ACT Thr	ATG Met	CGG	TCT	CCG	GTC	TTC	ACG	GAC	AAC	TCA	TCC	672
	210					215	501		vul	1116	220		ASII	Set	ser	
CCC	CCG	GCC	GTA	CCG	CAG	TCA	TTT	CAA	GTG	GCC	CAC	CTA	CAC	GCT	ccc	720
225	Pro	Ala	val	Pro	230	Ser	Phe	GIn	Val	A1a 235	His	Leu	His	Ala	Pro 240	
ACT	GGC Gly	AGC	GGC	AAG	AGT	ACT	AAA	GTG	CCG	GCT	GCA	TAT	GCA	GCC	CAA	768
				245					250					255		
GGG Gly	TAC Tyr	AAG Lvs	GTG Val	CTC Leu	GTC Val	CTC	AAT	CCG	TCC	GTT Val	GCC	GCT	ACC	TTA	GGG	816
			260					265					270		_	
TTT Phe	GGG Gly	GCG Ala	TAT	ATG Met	TCT	AAG Lvs	GCA Ala	CAC	GGT Glv	ATT	GAC	CCC	AAC	ATC	AGA	864
		275	_			-	280		2			285			9	
ACT	GGG	GTA	AGG	ACC	ATT	ACC	ACA	GGC	GCC	CCC	GTC	ACA	TAC	TCT	ACC	912
1111	Gly 290	vai	Arg	THE	TIE	295	Thr	GIY	AIA	Pro	300	Thr	Tyr	Ser	Thr	
TAT	GGC Gly	AAG	TTT	CTT	GCC	GAT	GGT	GGT	TGC	TCT	GGG	GGC	GCT	TAT	GAC	960
305		2,5	1110	Lea	310	nop	GIY	GIY	cys	315	GIY	GIĀ	Ala	Tyr	320	
	ATA Ile															1008
				325					330					335		
GGC Gly	ATC Ile	GGC Glv	ACA Thr	GTC Val	CTG Leu	GAC Asp	CAA Gln	GCG Ala	GAG Glu	ACG	GCT	GGA	GCG Als	CGG	CTT	1056
-	-		340					345	314	- * * *	- LLC	JLY	350	*zr d	neu	

Ome	a come															
Val	. Val	Lev 355	ı Ala	ACC Thr	GCT Ala	' ACG	Pro 360	Pro	GGA Gly	TCG	GTC Val	ACC Thr 369	· Val	CCA L Pro	CAC His	1104
CCA Pro	AAC Asn 370	: Ile	GAG Glu	GAG Glu	GTG Val	GCC Ala 375	Leu	TCT Ser	AAT Asn	ACT Thr	GGA G1y 380	glu	ATC	CCC Pro	TTC Phe	1152
TAT Tyr 385	GGC	AAA Lys	GCC Ala	ATC Ile	Pro 390	Ile	GAA Glu	GCC Ala	ATC Ile	AGG Arg 395	Gly	GGA Gly	AGG Arg	CAT His	CTC Leu 400	1200
ATT	TTC Phe	TGT Cys	CAT	TCC Ser 405	Lys	AAG Lys	AAG Lys	TGC Cys	GAC Asp 410	Glu	CTC Leu	GCC Ala	GCA Ala	AAG Lys 415	CTG Leu	1248
TCA Ser	GGC Gly	CTC	GGA Gly 420	ATC Ile	AAC Asn	GCT Ala	GTG Val	GCG Ala 425	Tyr	TAC Tyr	CGG Arg	GGG Gly	CTC Leu 430	Asp	GTG Val	1296
TCC Ser	GTC Val	ATA Ile 435	CCA Pro	ACT Thr	ATC Ile	GGA Gly	GAC Asp 440	GTC Val	GTT Val	GTC Val	GTG Val	GCA Ala 445	Thr	GAC Asp	GCT Ala	1344
CTG Leu	ATG Met 450	ACG Thr	GGC Gly	TAT Tyr	ACG Thr	GGC Gly 455	GAC Asp	TTT Phe	GAC Asp	TCA Ser	GTG Val 460	Ile	GAC Asp	TGT Cys	AAC Asn	1392
ACA Thr 465	TGT Cys	GTC Val	ACC Thr	CAG Gln	ACA Thr 470	GTC Val	GAC Asp	TTC Phe	AGC Ser	TTG Leu 475	GAT Asp	CCC Pro	ACC Thr	TTC Phe	ACC Thr 480	1440
ATT Ile	GAG Glu	ACG Thr	ACG Thr	ACC Thr 485	GTG Val	CCT Pro	CAA Gln	GAC Asp	GCA Ala 490	GTG Val	TCG Ser	CGC Arg	TCG Ser	CAG Gln 495	CGG Arg	1488
CGG Arg	GGT Gly	AGG Arg	ACT Thr 500	GGC Gly	AGG Arg	GGT Gly	AGG Arg	AGA Arg 505	GGC Gly	ATC Ile	TAC Tyr	AGG Arg	TTT Phe 510	GTG Val	ACT Thr	1536
CCG Pro	GGA Gly	GAA Glu 515	CGG Arg	CCC Pro	TCG Ser	GGC Gly	ATG Met 520	TTC Phe	GAT Asp	TCC Ser	TCG Ser	GTC Val 525	CTG Leu	TGT Cys	GAG Glu	1584
TGC Cys	TAT Tyr 530	GAC Asp	GCG Ala	GGC Gly	TGT Cys	GCT Ala 535	TGG Trp	TAC Tyr	GAG Glu	CTC Leu	ACC Thr 540	CCC Pro	GCC Ala	GAG Glu	ACC Thr	1632
TCG Ser 545	GTT Val	AGG Arg	TTG Leu	CGG Arg	GCC Ala 550	TAC Tyr	CTG Leu	AAC Asn	ACA Thr	CCA Pro 555	GGG Gly	TTG Leu	CCC Pro	GTT Val	TGC Cys 560	1680
CAG Gln	GAC Asp	CAC His	CTG Leu	GAG G1u 565	TTC Phe	TGG Trp	GAG Glu	AGT Ser	GTC Val 570	TTC Phe	ACA Thr	GGC Gly	CTC Leu	ACC Thr 575	CAT His	1728
ATA Ile	GAT Asp	GCA Ala	CAC His	TTC Phe	TTG Leu	TCC Ser	CAG . Gln	ACC Thr	AAG Lys	CAG G1n	GCA Ala	GGA Gly	GAC Asp	AAC Asn	TTC Phe	1776

580 585 CCC TAC CTG GTA GCA TAC CAA GCC ACG GTG TGC GCC AGG GCT CAG GCC 1824 Pro Tyr Leu Val Ala Tyr Gln Ala Thr Val Cys Ala Arg Ala Gln Ala 595 600 CCA CCT CCA TCA TGG GAT CAA ATG TGG AAG TGT CTC ATA CGG CTG AAA 1872 Pro Pro Pro Ser Trp Asp Gln Met Trp Lys Cys Leu Ile Arg Leu Lys 610 CCT ACG CTG CAC GGG CCA ACA CCC TTG CTG TAC AGG CTG GGA GCC GTC 1920 Pro Thr Leu His Gly Pro Thr Pro Leu Leu Tyr Arg Leu Gly Ala Val 630 635 1968 Gln Asn Glu Val Thr Leu Thr His Pro Ile Thr Lys Tyr Ile Met Ala 645 650 TGC ATG TCG GCT GAC CTG GAG GTC GTC ACT 1998 Cys Met Ser Ala Asp Leu Glu Val Val 660 (2) INFORMATION FOR SEQ ID NO:105: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 1998 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear (ii) MOLECULE TYPE: cDNA (ix) FEATURE: (A) NAME/KEY: CDS (B) LOCATION: 1..1995 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:105: ATG GGC AGC AGC CAT CAT CAT CAT CAC AGC AGC GGC CTG GTG CCG 48 Met Gly Ser Ser His His His His His Ser Ser Gly Leu Val Pro 5 10 CGC GGC AGC CAT ATG GGT TCT GTT GTT GTT GGT AGA ATT ATT TTA 96 Arg Gly Ser His Met Gly Ser Val Val Ile Val Gly Arg Ile Ile Leu TCT GGT AGT AGT ATC ACG GCC TAC TCC CAA CAG ACG CGG GGC CTA

Ser Gly Ser Gly Ser Ile Thr Ala Tyr Ser Gln Gln Thr Arg Gly Leu 35 40 45 CTT GGT TGC AAG AAG ACT AGC CTT ACA GGC CGG GAC AAG AAC CAG GTC

Leu Gly Cys Lys Lys Thr Ser Leu Thr Gly Arg Asp Lys Asn Gln Val

55

5.0

65 65	GIY	Glu	. Val	Gln	70	Val	Ser	Thr	: Ala	Thr 75	Gln	Ser	Phe	Let	GCG Ala 80		240
ACC Thr	TGC Cys	GTC Val	AAC Asn	GGC Gly 85	Val	TGT Cys	TGG	ACC Thr	GTT Val 90	Tyr	CAT His	GGT Gly	GCT Ala	GGC Gly 95	TCA Ser		288
AAG Lys	ACC Thr	TTA Leu	GCC Ala 100	Gly	CCA Pro	AAG Lys	GGG Gly	CCA Pro 105	Ile	ACC Thr	CAG Gln	ATG Met	TAC Tyr 110	Thr	AAT Asn		336
GTG Val	GAC Asp	CAG Gln 115	GAC Asp	CTC Leu	GTC Val	GGC Gly	TGG Trp 120	Gln	GCG Ala	CCC	CCC Pro	GGG Gly 125	Ala	CGT Arg	TCC	•	384
TTG Leu	ACA Thr 130	CCA Pro	TGC Cys	ACC Thr	TGT Cys	GGC Gly 135	AGC Ser	TCA Ser	GAC Asp	CTT Leu	TAC Tyr 140	TTG Leu	GTC Val	ACG Thr	AGA Arg		432
CAT His 145	GCT Ala	GAC Asp	GTC Val	ATT Ile	CCG Pro 150	GTG Val	CGC Arg	CGG Arg	CGG Arg	GGC Gly 155	GAC Asp	AGT Ser	AGG Arg	GGG Gly	AGC Ser 160		480
CTG Leu	CTC Leu	TCC Ser	CCC Pro	AGG Arg 165	CCT Pro	GTC Val	TCC Ser	TAC Tyr	TTG Leu 170	AAG Lys	GGC Gly	TCT Ser	TCG Ser	GGT Gly 175	GGT Gly		528
CCA Pro	CTG Leu	CTC Leu	TGC Cys 180	CCT Pro	TCG Ser	GGG Gly	CAC His	GCT Ala 185	GTG Val	GGC Gly	ATC Ile	TTC Phe	CGG Arg 190	GCT Ala	GCC Ala		576
GTA Val	TGC Cys	ACC Thr 195	CGG Arg	GGG Gly	GTT Val	GCG Ala	AAG Lys 200	GCG Ala	GTG Val	GAC Asp	TTT Phe	GTG Val 205	CCC Pro	GTA Val	GAG Glu		624
TCC Ser	ATG Met 210	GAA Glu	ACT Thr	ACT Thr	ATG Met	CGG Arg 215	TCT Ser	CCG Pro	GTC Val	TTC Phe	ACG Thr 220	GAC Asp	AAC Asn	TCA Ser	TCC Ser		672
CCC Pro 225	CCG Pro	GCC Ala	GTA Val	CCG Pro	CAG Gln 230	TCA Ser	TTT Phe	CAA Gln	GTG Val	GCC Ala 235	CAC His	CTA Leu	CAC His	GCT Ala	CCC Pro 240		720
ACT Thr	GGC Gly	AGC Ser	GGC Gly	AAG Lys 245	AGT Ser	ACT Thr	AAA Lys	GTG Val	CCG Pro 250	GCT Ala	GCA Ala	TAT Tyr	GCA Ala	GCC Ala 255	CAA Gln		768
GGG Gly	TAC Tyr	AAG Lys	GTG Val 260	CTC Leu	GTC Val	CTC Leu	AAT Asn	CCG Pro 265	TCC Ser	GTT Val	GCC Ala	GCT Ala	ACC Thr 270	TTA Leu	GGG Gly		816
TTT Phe	GGG Gly	GCG Ala 275	TAT Tyr	ATG Met	TCT Ser	Lys	GCA Ala 280	CAC His	GGT Gly	ATT Ile	GAC Asp	CCC Pro 285	AAC Asn	ATC Ile	AGA Arg		864
ACT Thr	GGG Gly	GTA Val	AGG Arg	ACC Thr	ATT Ile	ACC . Thr	ACA Thr	GGC Gly	GCC Ala	CCC Pro	GTC . Val	ACA Thr	TAC Tyr	TCT Ser	ACC Thr		912

290 295 300

	250					295	,				300)				
TAT Tyr 305	Gly	Lys	TTT Phe	CTI Leu	GCC Ala 310	Asp	GGT Gly	GGT Gly	TGC Cys	TCT Ser 315	Gly	GGC Gly	GCT Ala	TAT Tyr	GAC Asp 320	960
ATC Ile	ATA Ile	ATA	TGT Cys	GAT Asp 325	Glu	TGC Cys	CAT His	TCA Ser	ACT Thr 330	Asp	TCG Ser	ACT Thr	ACA Thr	ATC Ile 335	TTG Leu	1008
GGC Gly	ATC Ile	GGC Gly	ACA Thr 340	Val	CTG Leu	GAC Asp	CAA Gln	GCG Ala 345	Glu	ACG Thr	GCT Ala	GGA Gly	GCG Ala 350	Arg	CTT Leu	1056
Val	Val	Leu 355		Thr	Ala	Thr	Pro 360	Pro	Gly	Ser	Val	Thr 365	Val	Pro	His	1104
Pro	Asn 370	Ile	GAG Glu	Glu	Val	Ala 375	Leu	Ser	Asn	Thr	Gly 380	Glu	Ile	Pro	Phe	1152
Tyr 385	Gly	Lys	GCC Ala	Ile	Pro 390	Ile	Glu	Ala	Ile	Arg 395	Gly	Gly	Arg	His	Leu 400	1200
Ile	Phe	Cys	CAT	Ser 405	Lys	Lys	Lys	Cys	Asp 410	Glu	Leu	Ala	Ala	Lys 415	Leu	1248
Ser	Gly	Leu	GGA Gly 420	Ile	Asn	Ala	Val	Ala 425	Tyr	Tyr	Arg	Gly	Leu 430	Asp	Val	1296
Ser	Val	11e 435	CCA Pro	Thr	Ile	Gly	Asp 440	Val	Val	Va1	Val	Ala 445	Thr	Asp	Ala	1344
Leu	Met 450	Thr	GGC Gly	Tyr	Thr	Gly 455	Asp	Phe	Asp	Ser	Val 460	Ile	Asp	Cys	Asn	1392
ACA Thr 465	TGT Cys	GTC Val	ACC Thr	CAG Gln	ACA Thr 470	GTC Val	GAC Asp	TTC Phe	AGC Ser	TTG Leu 475	GAT Asp	CCC Pro	ACC Thr	TTC Phe	ACC Thr 480	1440
ATT	GAG Glu	ACG Thr	ACG Thr	ACC Thr 485	GTG Val	CCT Pro	CAA Gln	GAC Asp	GCA Ala 490	GTG Val	TCG Ser	CGC Arg	TCG Ser	CAG Gln 495	CGG Arg	1488
Arg	Gly	Arg	ACT Thr 500	Gly	Arg	Gly	Arg	Arg 505	Gly	Ile	Tyr	Arg	Phe 510	Val	Thr	1536
CCG Pro	GGA Gly	GAA Glu 515	CGG Arg	CCC Pro	TCG Ser	GGC Gly	ATG Met 520	TTC Phe	GAT Asp	TCC Ser	TCG Ser	GTC Val 525	CTG Leu	TGT Cys	GAG Glu	1584

TGC Cys	TAT Tyr 530	Asp	GCG Ala	GGC Gly	TGT	GCT Ala 535	Trp	TAC	GAG Glu	CTC Leu	ACC Thr 540	Pro	GCC Ala	GAG Glu	ACC Thr	163
TCG Ser 545	GTT Val	AGG Arg	TTG Leu	CGG Arg	GCC Ala 550	TAC Tyr	CTG Leu	AAC Asn	ACA Thr	CCA Pro 555	GGG Gly	TTG Leu	CCC Pro	GTT Val	TGC Cys 560	1680
CAG Gln	GAC Asp	CAC	CTG Leu	GAG Glu 565	TTC Phe	TGG Trp	GAG Glu	AGT Ser	GTC Val 570	TTC Phe	ACA Thr	GGC Gly	CTC Leu	ACC Thr 575	His	1728
ATA Ile	GAT Asp	GCA Ala	CAC His 580	TTC Phe	TTG Leu	TCC Ser	CAG Gln	ACC Thr 585	AAG Lys	CAG Gln	GCA Ala	GGA Gly	GAC Asp 590	AAC Asn	TTC Phe	1776
CCC Pro	TAC Tyr	CTG Leu 595	GTA Val	GCA Ala	TAC Tyr	CAA Gln	GCC Ala 600	ACG Thr	GTG Val	TGC Cys	GCC Ala	AGG Arg 605	GCT Ala	CAG Gln	GCC Ala	1824
CCA Pro	CCT Pro 610	CCA Pro	TCA Ser	TGG Trp	GAT Asp	CAA Gln 615	ATG Met	TGG Trp	AAG Lys	TGT Cys	CTC Leu 620	ATA Ile	CGG Arg	CTG Leu	AAA Lys	1872
CCT Pro 625	ACG Thr	CTG Leu	CAC His	GGG Gly	CCA Pro 630	ACA Thr	CCC Pro	TTG Leu	CTG Leu	TAC Tyr 635	AGG Arg	CTG Leu	GGA Gly	GCC Ala	GTC Val 640	1920
CAA Gln	AAT Asn	GAG Glu	GTC Val	ACC Thr 645	CTC Leu	ACC Thr	CAC His	CCC Pro	ATA Ile 650	ACC Thr	AAA Lys	TAC Tyr	ATC Ile	ATG Met 655	GCA Ala	1968
			GCT Ala 660						ACT							1998

- (2) INFORMATION FOR SEQ ID NO:106:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1998 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: cDNA
 - (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION: 1..1995
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:106:

				Met					Ile			AGA Arg		Ile	TTA Leu	96
TCT Ser	GGT Gly	AGT Ser 35	Gly	AGT Ser	ATC Ile	ACG Thr	GCC Ala 40	Tyr	TCC Ser	CAA G1n	CAG Gln	ACG Thr 45	Arg	GGC Gly	CTA Leu	144
CTT Leu	GGT Gly 50	TGC Cys	ATC Ile	ATC Ile	ACT Thr	AGC Ser 55	CTT Leu	ACA Thr	GGC Gly	CGG Arg	GAC Asp 60	AAG Lys	AAC Asn	CAG Gln	GTC Val	192
GAG Glu 65	GGA Gly	GAG Glu	GTT Val	CAG Gln	GTG Val 70	GTT Val	TCC Ser	ACC Thr	GCA Ala	ACA Thr 75	CAA Gln	TCC Ser	TTC Phe	CTG Leu	GCG Ala 80	240
												GGT Gly				288
AAG Lys	ACC Thr	TTA Leu	GCC Ala 100	GGC Gly	CCA Pro	AAG Lys	GGG Gly	CCA Pro 105	ATC Ile	ACC Thr	CAG Gln	ATG Met	TAC Tyr 110	ACT Thr	AAT Asn	336
												GGG Gly 125				384
TTG Leu	ACA Thr 130	CCA Pro	TGC Cys	ACC Thr	TGT Cys	GGC Gly 135	AGC Ser	TCA Ser	GAC Asp	CTT Leu	TAC Tyr 140	TTG Leu	GTC Val	ACG Thr	AGA Arg	432
CAT His 145	GCT Ala	GAC Asp	GTC Val	ATT Ile	CCG Pro 150	GTG Val	CGC Arg	CGG Arg	CGG Arg	GGC Gly 155	GAC Asp	AGT Ser	AGG Arg	GGG Gly	AGC Ser 160	480
												TCT Ser				528
CCA Pro	CTG Leu	CTC Leu	TGC Cys 180	CCT Pro	TCG Ser	GGG Gly	CAC His	GCT Ala 185	GTG Val	GGC Gly	ATC Ile	TTC Phe	CGG Arg 190	GCT Ala	GCC Ala	576
GTA Val	TGC Cys	ACC Thr 195	CGG Arg	GGG Gly	GTT Val	GCG Ala	AAG Lys 200	GCG Ala	GTG Val	GAC Asp	TTT Phe	GTG Val 205	CCC Pro	GTA Val	GAG Glu	624
TCC Ser	ATG Met 210	GAA Glu	ACT Thr	ACT Thr	ATG Met	CGG Arg 215	TCT Ser	CCG Pro	GTC Val	TTC Phe	ACG Thr 220	GAC Asp	AAC Asn	TCA Ser	TCC Ser	672
CCC Pro 225	CCG Pro	GCC Ala	GTA Val	CCG Pro	CAG Gln 230	TCA Ser	TTT Phe	CAA Gln	GTG Val	GCC Ala 235	CAC His	CTA Leu	CAC His	GCT Ala	CCC Pro 240	720
ACT	GGC	AGC	GGC	AAG	AGT	ACT	AAA	GTG	CCG	GCT	GCA	TAT	GCA	GCC	CAA	768

Thr	Gly	Ser	Gly	Lys 245	Ser	Thr	Lys	Val	250		Ala	a Tyr	Ala	Ala 255	Gln	
GGG Gly	TAC	AAG Lys	GTG Val 260	Leu	GTC Val	CTC Leu	AAT Asn	CCG Pro 265	Ser	GTT Val	GCC Ala	GCT Ala	ACC Thr 270	Leu	GGG Gly	816
TTT Phe	GGG Gly	GCG Ala 275	Tyr	ATG Met	TCT	AAG Lys	GCA Ala 280	His	GGT Gly	ATT Ile	GAC Asp	CCC Pro 285	Asn	ATC Ile	AGA Arg	864
ACT Thr	GGG Gly 290	GTA Val	AGG Arg	ACC Thr	ATT	ACC Thr 295	Thr	GGC Gly	GCC Ala	CCC Pro	GTC Val 300	ACA Thr	TAC Tyr	TCT Ser	ACC Thr	912
TAT Tyr 305	GGC Gly	AAG Lys	TTT Phe	CTT Leu	GCC Ala 310	GAT Asp	GGT Gly	GGT Gly	TGC Cys	TCT Ser 315	GGG Gly	GGC Gly	GCT Ala	TAT Tyr	GAC Asp 320	960
ATC Ile	ATA Ile	ATA Ile	TGT Cys	GAT Asp 325	GAG Glu	TGC Cys	CAT His	TCA Ser	ACT Thr 330	GAC Asp	TCG Ser	ACT Thr	ACA Thr	ATC Ile 335	TTG Leu	1008
GGC Gly	ATC Ile	GGC Gly	ACA Thr 340	GTC Val	CTG Leu	GAC Asp	CAA Gln	GCG Ala 345	GAG Glu	ACG Thr	GCT Ala	GGA Gly	GCG Ala 350	CGG Arg	CTT Leu	1056
GTC Val	GTG Val	CTC Leu 355	GCC Ala	ACC Thr	GCT Ala	ACG Thr	CCT Pro 360	CCG Pro	GGA Gly	TCG Ser	GTC Val	ACC Thr 365	GTG Val	CCA Pro	CAC His	1104
CCA Pro	AAC Asn 370	ATC Ile	GAG Glu	GAG Glu	GTG Val	GCC Ala 375	CTG Leu	TCT Ser	AAT Asn	ACT Thr	GGA Gly 380	GAG Glu	ATC Ile	CCC Pro	TTC Phe	1152
TAT Tyr 385	GGC Gly	AAA Lys	GCC Ala	ATC Ile	CCC Pro 390	ATT Ile	GAA Glu	GCC Ala	ATC Ile	AGG Arg 395	GGG Gly	GGA Gly	AGG Arg	CAT His	CTC Leu 400	1200
ATT Ile	TTC Phe	TGT Cys	CAT His	TCC Ser 405	AAG Lys	AAG Lys	AAG Lys	TGC Cys	GAC Asp 410	GAG Glu	CTC Leu	GCC Ala	GCA Ala	AAG Lys 415	CTG Leu	1248
TCA Ser	GGC Gly	CTC Leu	GGA Gly 420	ATC Ile	AAC Asn	GCT Ala	GTG Val	GCG A1a 425	TAT Tyr	TAC Tyr	CGG Arg	GGG Gly	CTC Leu 430	GAT Asp	GTG Val	1296
TCC Ser	GTC Val	ATA Ile 435	CCA Pro	ACT Thr	ATC Ile	GGA Gly	GAC Asp 440	GTC Val	GTT Val	GTC Val	GTG Val	GCA Ala 445	ACA Thr	GAC Asp	GCT Ala	1344
CTG Leu	ATG Met 450	ACG Thr	GGC Gly	TAT Tyr	ACG Thr	GGC G1y 455	GAC Asp	TTT Phe	GAC Asp	TCA Ser	GTG Val 460	ATC Ile	GAC Asp	TGT Cys	AAC Asn	1392
ACA Thr 465	TGT Cys	GTC Val	ACC Thr	CAG Gln	ACA Thr 470	GTC Val	GAC Asp	TTC Phe	AGC Ser	TTG Leu 475	GAT Asp	CCC Pro	ACC Thr	TTC Phe	ACC Thr 480	1440

ATT	GAG Glu	ACG Thr	ACG	Thr 485	Val	CCT Pro	CAA Glr	GAC Asp	GCA Ala 490	Val	TCG Ser	CGC Arg	TCG Ser	CAG Gln 495	Arg	1488
CGG Arg	GGT Gly	AGG Arg	ACT Thr 500	Gly	AGG Arg	GGT Gly	AGG Arg	AGA Arg 505	Gly	ATC Ile	TAC Tyr	AGG Arg	TTT Phe 510	Val	ACT Thr	1536
CCG Pro	GGA Gly	GAA Glu 515	Arg	CCC	TCG Ser	GGC Gly	ATG Met 520	Phe	GAT Asp	TCC Ser	TCG Ser	GTC Val 525	Leu	TGT Cys	GAG Glu	1584
TGC Cys	TAT Tyr 530	GAC Asp	GCG Ala	GGC Gly	TGT Cys	GCT Ala 535	TGG Trp	TAC Tyr	GAG Glu	CTC Leu	ACC Thr 540	Pro	GCC Ala	GAG Glu	ACC Thr	1632
TCG Ser 545	GTT Val	AGG Arg	TTG Leu	CGG Arg	GCC Ala 550	TAC Tyr	CTG Leu	AAC Asn	ACA Thr	CCA Pro 555	GGG Gly	TTG Leu	CCC Pro	GTT Val	TGC Cys 560	1680
CAG Gln	GAC Asp	CAC His	CTG Leu	GAG Glu 565	TTC Phe	TGG Trp	GAG Glu	AGT Ser	GTC Val 570	TTC Phe	ACA Thr	GGC Gly	CTC Leu	ACC Thr 575	CAT His	1728
ATA Ile	GAT Asp	GCA Ala	CAC His 580	TTC Phe	TTG Leu	TCC Ser	CAG Gln	ACC Thr 585	AAG Lys	CAG Gln	GCA Ala	GGA Gly	GAC Asp 590	AAC Asn	TTC Phe	1776
CCC Pro	TAC Tyr	CTG Leu 595	GTA Val	GCA Ala	TAC Tyr	CAA Gln	GCC Ala 600	ACG Thr	GTG Val	TGC Cys	GCC Ala	AGG Arg 605	GCT Ala	CAG Gln	GCC Ala	1824
CCA Pro	CCT Pro 610	CCA Pro	TCA Ser	TGG Trp	GAT Asp	CAA Gln 615	ATG Met	TGG Trp	AAG Lys	TGT Cys	CTC Leu 620	ATA Ile	CGG Arg	CTG Leu	AAA Lys	1872
CCT Pro 625	ACG Thr	CTG Leu	CAC His	GGG Gly	CCA Pro 630	ACA Thr	CCC Pro	TTG Leu	CTG Leu	TAC Tyr 635	AGG Arg	CTG Leu	GGA Gly	GCC Ala	GTC Val 640	1920
CAA Gln	AAT Asn	GAG Glu	GTC Val	ACC Thr 645	CTC Leu	ACC Thr	CAC His	CCC Pro	ATA Ile 650	ACC Thr	AAA Lys	TAC Tyr	ATC Ile	ATG Met 655	GCA Ala	1968
	ATG Met								ACT							1998

(2) INFORMATION FOR SEQ ID NO:107:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1998 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 1..1997

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 1..1995

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:107:

ATG Met 1	GGC Gly	AGC Ser	AGC Ser	CAT His 5	His	CAT His	CAT His	CAT His	CAC His 10	AGC Ser	AGC Ser	GGC Gly	CTG Leu	GTG Val 15	Pro	48
CGC Arg	GGC Gly	AGC Ser	CAT His 20	ATG Met	GGT Gly	TCT	GTT Val	GTT Val 25	ATT	GTT Val	GGT Gly	AGA Arg	ATT Ile 30	ATT Ile	TTA. Leu	96
	GGT Gly												Arg			144
	GGT Gly 50											Lys				192
	GGA Gly															240
	TGC Cys															288
	ACC Thr															336
	GAC Asp															384
	ACA Thr 130															432
	GCT Ala															480
	CTC Leu															528

CCA Pro	CTG Let	CTC	TGC Cys 180	Pro	TCG Ser	GGG Gly	CAC His	GCT Ala 185	. Val	GGC Gly	ATC	TTC Phe	CGG Arg	Ala	GCC Ala	576
GTA Val	TGC Cys	Thr	Arg	GGG Gly	GTT Val	GCG Ala	AAG Lys 200	Ala	GTG Val	GAC Asp	TTT Phe	GTG Val 205	Pro	GTA Val	GAG Glu	624
TCC Ser	ATG Met 210	Glu	ACT Thr	ACT Thr	ATG Met	CGG Arg 215	Ser	CCG Pro	GTC Val	TTC Phe	ACG Thr 220	Asp	AAC Asn	TCA Ser	TCC Ser	672
CCC Pro 225	CCG Pro	GCC Ala	GTA Val	CCG Pro	CAG Gln 230	Ser	TTT Phe	CAA Gln	GTG Val	GCC Ala 235	His	CTA Leu	CAC His	GCT Ala	CCC Pro 240	. 720
ACT Thr	GGC Gly	AGC Ser	GGC Gly	AAG Lys 245	AGT Ser	ACT Thr	AAA Lys	GTG Val	CCG Pro 250	GCT Ala	GCA Ala	TAT Tyr	GCA Ala	GCC Ala 255	Gln	768
GGG Gly	TAC Tyr	AAG Lys	GTG Val 260	CTC Leu	GTC Val	CTC Leu	AAT Asn	CCG Pro 265	TCC Ser	GTT Val	GCC Ala	GCT Ala	ACC Thr 270	TTA Leu	GGG Gly	816
TTT Phe	GGG Gly	GCG Ala 275	TAT Tyr	ATG Met	TCT Ser	AAG Lys	GCA Ala 280	CAC His	GGT Gly	ATT Ile	GAC Asp	CCC Pro 285	AAC Asn	ATC Ile	AGA Arg	864
ACT Thr	GGG Gly 290	GTA Val	AGG Arg	ACC Thr	ATT Ile	ACC Thr 295	ACA Thr	GGC Gly	GCC Ala	CCC Pro	GTC Val 300	Thr	TAC Tyr	TCT Ser	ACC Thr	912
TAT Tyr 305	GGC Gly	AAG Lys	TTT Phe	CTT Leu	GCC Ala 310	GAT Asp	GGT Gly	GGT Gly	TGC Cys	TCT Ser 315	GGG Gly	GGC Gly	GCT Ala	TAT Tyr	GAC Asp 320	960
ATC Ile	ATA Ile	ATA Ile	TGT Cys	GAT Asp 325	GAG Glu	TGC Cys	CAT His	TCA Ser	ACT Thr 330	GAC Asp	TCG Ser	ACT Thr	ACA Thr	ATC Ile 335	TTG Leu	1008
GGC Gly	ATC Ile	GGC Gly	ACA Thr 340	GTC Val	CTG Leu	GAC Asp	CAA Gln	GCG Ala 345	GAG Glu	ACG Thr	GCT Ala	GGA Gly	GCG Ala 350	CGG Arg	CTT Leu	1056
GTC Val	GTG Val	CTC Leu 355	GCC Ala	ACC Thr	GCT Ala	ACG Thr	CCT Pro 360	CCG Pro	GGA Gly	TCG Ser	GTC Val	ACC Thr 365	GTG Val	CCA Pro	CAC His	1104
CCA Pro	AAC Asn 370	ATC Ile	GAG Glu	GAG Glu	GTG Val	GCC Ala 375	CTG Leu	TCT Ser	AAT Asn	ACT Thr	GGA Gly 380	GAG Glu	ATC Ile	CCC Pro	TTC Phe	1152
TAT Tyr 385	GGC G1y	AAA Lys	GCC Ala	ATC Ile	CCC Pro 390	ATT Ile	GAA Glu	GCC Ala	ATC Ile	AGG Arg 395	GGG Gly	GGA Gly	AGG Arg	CAT His	CTC Leu 400	1200
ATT Ile	TTC Phe	TGT Cys	CAT His	TCC Ser	AAG Lys	AAG Lys	AAG Lys	TGC Cys	GAC Asp	GAG Glu	CTC Leu	GCC Ala	GCA Ala	AAG Lys	CTG Leu	1248

405 410 415 TCA GGC CTC GGA ATC AAC GCT GTG GCG TAT TAC CGG GGG CTC GAT GTG 1296 Ser Gly Leu Gly Ile Asn Ala Val Ala Tyr Tyr Arg Gly Leu Asp Val 420 425 TCC GTC ATA CCA ACT ATC GGA GAC GTC GTT GTC GTG GCA ACA GAC GCT 1344 Ser Val Ile Pro Thr Ile Gly Asp Val Val Val Val Ala Thr Asp Ala 435 440 CTG ATG ACG GGC TAT ACG GGC GAC TTT GAC TCA GTG ATC GAC TGT AAC 1392 Leu Met Thr Gly Tyr Thr Gly Asp Phe Asp Ser Val Ile Asp Cys Asn 450 455 ACA TGT GTC ACC CAG ACA GTC GAC TTC AGC TTG GAT CCC ACC TTC ACC 1440 Thr Cys Val Thr Gln Thr Val Asp Phe Ser Leu Asp Pro Thr Phe Thr 470 ATT GAG ACG ACG GTG CCT CAA GAC GCA GTG TCG CGC TCG CAG CGG 1488 Ile Glu Thr Thr Thr Val Pro Gln Asp Ala Val Ser Arg Ser Gln Arg 485 490 CGG GGT AGG ACT GGC AGG GGT AGG AGA GGC ATC TAC AGG TTT GTG ACT 1536 Arg Gly Arg Thr Gly Arg Gly Arg Gly Ile Tyr Arg Phe Val Thr 500 505 CCG GGA GAA CGG CCC TCG GGC ATG TTC GAT TCC TCG GTC CTG TGT GAG 1584 Pro Gly Glu Arg Pro Ser Gly Met Phe Asp Ser Ser Val Leu Cys Glu 515 520 TGC TAT GAC GCG GGC TGT GCT TGG TAC GAG CTC ACC CCC GCC GAG ACC 1632 Cys Tyr Asp Ala Gly Cys Ala Trp Tyr Glu Leu Thr Pro Ala Glu Thr 530 TCG GTT AGG TTG CGG GCC TAC CTG AAC ACA CCA GGG TTG CCC GTT TGC 1680 Ser Val Arg Leu Arg Ala Tyr Leu Asn Thr Pro Gly Leu Pro Val Cys 545 550 555 CAG GAC CAC CTG GAG TTC TGG GAG AGT GTC TTC ACA GGC CTC ACC CAT 1728 Gln Asp His Leu Glu Phe Trp Glu Ser Val Phe Thr Gly Leu Thr His 565 ATA GAT GCA CAC TTC TTG TCC CAG ACC AAG CAG GCA GGA GAC AAC TTC 1776 Ile Asp Ala His Phe Leu Ser Gln Thr Lys Gln Ala Gly Asp Asn Phe 580 585 CCC TAC CTG GTA GCA TAC CAA GCC ACG GTG TGC GCC AGG GCT CAG GCC 1824 Pro Tyr Leu Val Ala Tyr Gln Ala Thr Val Cys Ala Arg Ala Gln Ala

CCA CCT CCA TCA TGG GAT CAA ATG TGG AAG TGT CTC ATA CGG CTG AAA
Pro Pro Pro Ser Trp Asp Gln Met Trp Lys Cys Leu Ile Arg Leu Lys
610 620

CCT ACG CTG CAC GGG CCA ACA CCC TTG CTG TAC AGG CTG GGA GCC GTC 1920

Pro Thr Leu His Gly Pro Thr Pro Leu Leu Tyr Arg Leu Gly Ala Val 625 630 635 640

600

CAA Gln	AAT Asn	GAG Glu	GTC Val	ACC Thr 645	Leu	ACC Thr	CAC	CCC Pro	ATA Ile	Thr	AAA Lys	TAC Tyr	ATC Ile	ATG Met	GCA : Ala	196
				Asp	CTG Leu				L							1998
(2)	INF	ORMA	TION	FOR	SEQ	ID I	NO:1	08:								
	(1	1	(A) I (B) 5 (C) 5	LENGT TYPE : STRAI	CHARA CH: 1 : nuc NDEDN LOGY:	L998 Cleic NESS:	base ac: sir	e pai id	rs							
	(ii	.) MC	LECU	JLE T	YPE:	cDi	VA.								٠	
	(i>		(A)	IAME/	KEY:			5								
	(xi) SE	QUEN	CE D	ESCR	IPTI	ON:	SEQ	ID N	0:10	18					
ATG Met 1	GGC Gly	AGC Ser	AGC Ser	CAT His 5	CAT His	CAT	CAT His	CAT His	CAC His 10	AGC Ser	AGC Ser	GGC Gly	CTG Leu	GTG Val 15	CCG Pro	4.8
CGC Arg	GGC Gly	AGC Ser	CAT His 20	ATG Met	GGT Gly	TCT Ser	GTT Val	GTT Val 25	ATT Ile	GTT Val	GGT Gly	AGA Arg	ATT Ile 30	Ile	TTA Leu	96
TCT Ser	GGT Gly	AGT Ser 35	GGT Gly	AGT Ser	ATC Ile	ACG Thr	GCC Ala 40	TAC Tyr	TCC Ser	CAA Gln	CAG Gln	ACG Thr 45	CGG Arg	GGC Gly	CTA Leu	144
CTT Leu	GGT Gly 50	TGC Cys	ATC Ile	AAG Lys	ACT Thr	AGC Ser 55	CTT Leu	ACA Thr	GGC Gly	CGG Arg	GAC Asp 60	AAG Lys	AAC Asn	CAG Gln	GTC Val	192
GAG Glu 65	GGA Gly	GAG Glu	GTT Val	CAG Gln	GTG Val 70	GTT Val	TCC Ser	ACC Thr	GCA Ala	ACA Thr 75	CAA Gln	TCC Ser	TTC Phe	CTG Leu	GCG Ala 80	240
					GTG Val											288
					CCA Pro											336
GTG	GAC	CAG	GAC	CTC	стс	GGC	TICC	CAC	ccc	ccc	ccc	ccc	ccc	CCM	mcc.	204

Val	. Asp	Gln 115	Asp	Leu	Val	Gly	120		Ala	Pro	Pro	Gly 125		Arg	Ser	
TTG Leu	ACA Thr 130	Pro	TGC Cys	ACC	TGT Cys	GGC Gly 135	Ser	TCA Ser	GAC Asp	CTT Leu	TAC Tyr 140	Leu	GTC Val	ACG Thr	AGA Arg	432
CAT His 145	GCT Ala	GAC Asp	GTC Val	ATT	CCG Pro 150	Val	CGC Arg	CGG Arg	CGG	GGC Gly 155	Asp	AGT Ser	AGG Arg	GGG Gly	AGC Ser 160	480
CTG Leu	CTC Leu	TCC Ser	CCC	AGG Arg 165	Pro	GTC Val	TCC Ser	TAC Tyr	TTG Leu 170	Lys	GGC Gly	TCT	GCT Ala	GGT Gly 175	Gly	528
CCA Pro	CTG Leu	CTC Leu	TGC Cys 180	CCT Pro	TCG Ser	GGG Gly	CAC His	GCT Ala 185	Val	GGC Gly	ATC Ile	TTC Phe	CGG Arg 190	Ala	GCC Ala	576
GTA Val	TGC Cys	ACC Thr 195	CGG Arg	GGG Gly	GTT Val	GCG Ala	AAG Lys 200	Ala	GTG Val	GAC Asp	TTT Phe	GTG Val 205	Pro	GTA Val	GAG Glu	624
TCC Ser	ATG Met 210	GAA Glu	ACT Thr	ACT Thr	ATG Met	CGG Arg 215	TCT Ser	CCG Pro	GTC Val	TTC Phe	ACG Thr 220	GAC Asp	AAC Asn	TCA Ser	TCC Ser	672
CCC Pro 225	CCG Pro	GCC Ala	GTA Val	CCG Pro	CAG Gln 230	TCA Ser	TTT Phe	CAA Gln	GTG Val	GCC Ala 235	CAC His	CTA Leu	CAC His	GCT Ala	CCC Pro 240	720
ACT Thr	GGC Gly	AGC Ser	GGC Gly	AAG Lys 245	AGT Ser	ACT Thr	AAA Lys	GTG Val	CCG Pro 250	GCT Ala	GCA Ala	TAT Tyr	GCA Ala	GCC Ala 255	CAA Gln	768
GGG Gly	TAC Tyr	AAG Lys	GTG Val 260	CTC Leu	GTC Val	CTC Leu	AAT Asn	CCG Pro 265	TCC Ser	GTT Val	GCC Ala	GCT Ala	ACC Thr 270	TTA Leu	GGG Gly	816
TTT Phe	GGG Gly	GCG Ala 275	TAT Tyr	ATG Met	TCT Ser	AAG Lys	GCA Ala 280	CAC His	GGT Gly	ATT Ile	GAC Asp	CCC Pro 285	AAC Asn	ATC Ile	AGA Arg	864
ACT Thr	GGG Gly 290	GTA Val	AGG Arg	ACC Thr	ATT Ile	ACC Thr 295	ACA Thr	GGC Gly	GCC Ala	CCC Pro	GTC Val 300	ACA Thr	TAC Tyr	TCT Ser	ACC Thr	912
TAT Tyr 305	GGC Gly	AAG Lys	TTT Phe	CTT Leu	GCC Ala 310	GAT Asp	GGT Gly	GGT Gly	TGC Cys	TCT Ser 315	GGG Gly	GGC Gly	GCT Ala	TAT Tyr	GAC Asp 320	960
	ATA Ile															1008
GGC Gly	ATC Ile	Gly	ACA Thr 340	GTC Val	CTG Leu	GAC Asp	CAA Gln	GCG Ala 345	GAG Glu	ACG Thr	GCT Ala	GGA Gly	GCG Ala 350	CGG Arg	CTT Leu	1056

		CTC Leu 355											Val			1104
		ATC Ile										Glu				1152
		AAA Lys														1200
		TGT Cys														1248
		CTC Leu														1296
		ATA Ile 435														1344
		ACG Thr										Ile				1392
		GTC Val														1440
		ACG Thr														1488
		AGG Arg														1536
		GAA Glu 515														1584
		GAC Asp														1632
		AGG Arg														1680
		CAC His														1728
ATA	GAT	GCA	CAC	TTC	TTG	TCC	CAG	ACC	AAG	CAG	GCA	GGA	GAC	AAC	TTC	1776

11e	Asp	Ala	H1S 580	Phe	Leu	Ser	Gln	Thr 585	Lys	Gln	Ala	Gly	Asp 590	Asn	Phe	
CCC	TAC Tyr	CTG Leu 595	GTA Val	GCA Ala	TAC Tyr	CAA Gln	GCC Ala 600	ACG Thr	GTG Val	TGC Cys	GCC Ala	AGG Arg 605	GCT Ala	CAG Gln	GCC Ala	1824
CCA Pro	CCT Pro 610	CCA Pro	TCA Ser	TGG Trp	GAT Asp	CAA Gln 615	ATG Met	TGG Trp	AAG Lys	TGT Cys	CTC Leu 620	Ile	CGG Arg	CTG Leu	AAA Lys	1872
	ACG Thr														GTC Val 640	1920
	AAT Asn															1968
	ATG Met															1998
(2)	(ii (ix (ix) SE (. (. (. (. (. (. (. (. (. (. (. (. (.	QUEN A) L B) T C) S D) T LECU ATUR A) N B) L ATUR A) N B) L	CE CE CE ENGT YPE: TRAN OPOL LE T E: AME/OCAT E: AME/OCAT	HARA H: 1 nuc DEDN OGY: YPE: KEY: ION: KEY:	CTER 998 leic ESS: lin CDN CDS 1	ISTI base aci sin ear A	CS: pai d gle		0:10	9:					
	GGC Gly															48
	GGC Gly															96
TCT Ser	GGT Gly	AGT Ser 35	GGT Gly	AGT Ser	ATC Ile	ACG Thr	GCC Ala 40	TAC Tyr	TCC Ser	CAA Gln	CAG Gln	ACG Thr 45	CGG Arg	GGC Gly	CTA Leu	144

															GTC Val	192
	50					55			Ī		60	-				
GAG Glu 65	GGA Gly	GAG Glu	GTT Val	CAG Gln	GTG Val 70	GTT Val	TCC	ACC Thr	GCA Ala	ACA Thr 75	CAA Gln	TCC Ser	TTC Phe	CTG Leu	GCG Ala 80	240
			AAC Asn													288
AAG Lys	ACC Thr	TTA Leu	GCC Ala 100	GGC Gly	CCA Pro	AAG Lys	GGG Gly	CCA Pro 105	Ile	ACC Thr	CAG Gln	ATG Met	TAC Tyr 110	Thr	AAT Asn	336
			GAC Asp													384
			TGC Cys													432
			GTC Val													480
			CCC Pro													528
CCA Pro	CTG Leu	CTC Leu	TGC Cys 180	CCT Pro	TCG Ser	GGG Gly	CAC His	GCT Ala 185	GTG Val	GGC Gly	ATC Ile	TTC Phe	CGG Arg 190	GCT Ala	GCC Ala	576
			CGG Arg													624
			ACT Thr													672
			GTA Val													720
			GGC Gly													768
GGG Gly	TAC Tyr	AAG Lys	GTG Val 260	CTC Leu	GTC Val	CTC Leu	AAT Asn	CCG Pro 265	TCC Ser	GTT Val	GCC Ala	GCT Ala	ACC Thr 270	TTA Leu	GGG Gly	816
TTT Phe			TAT Tyr													864

	275			280			285		
GGG Gly 290				Thr			Thr		912
GGC Gly									960
ATA Ile									1008
ATC Ile									1056
GTG Val									1104
AAC Asn 370							Glu		1152
GGC Gly									1200
TTC Phe									1248
GGC Gly									1296
GTC Val									1344
ATG Met 450									1392
TGT Cys									1440
GAG Glu									1488
GGT Gly									1536

TOIGETEE ILELOS

CCG	GGA	GAA	CGG	CCC	TCG	GGC	ATG	TTC	GAT	TCC	TCG	GTC	CTG	TGT	GAG	158
PIO	Gly	515	Arg	Pro	Ser	. GIĀ	520		Asp	Ser	Ser	Val 525		Cys	Glu	
TGC	TAT	GAC	GCG	GGC	TGT	GCT	TGG	TAC	GAG	CTC	ACC	CCC	GCC	GAG	ACC	1632
Cys	Tyr 530	ASP	ATA	GIĀ	Cys	535	Trp	Tyr	Glu	Leu	Thr 540		Ala	Glu	Thr	
TCG	GTT	AGG	TTG	CGG	GCC	TAC	CTG	AAC	ACA	CCA	GGG	TTG	CCC	GTT	TGC	1680
545	Val	Arg	Leu	Arg	550		Leu	Asn	Thr	Pro 555	Gly	Leu	Pro	Val	Cys 560	
CAG	GAC	CAC	CTG	GAG	TTC	TGG	GAG	AGT	GTC	TTC	ACA	GGC	CTC	ACC	CAT	1728
GIn	Asp	His	Leu	Glu 565	Phe	Trp	Glu	Ser	Val 570	Phe	Thr	G1y	Leu	Thr 575	His	
ATA	GAT	GCA	CAC	TTC	TTG	TCC	CAG	ACC	AAG	CAG	GCA	GGA	GAC	AAC	TTC	1776
ITe	Asp	Ala	His 580	Phe	Leu	Ser	Gln	Thr 585	Lys	Gln	Ala	Gly	Asp 590	Asn	Phe	
CCC	TAC	CTG	GTA	GCA	TAC	CAA	GCC	ACG	GTG	TGC	GCC	AGG	GCT	CAG	GCC	1824
Pro	Tyr	Leu 595	Val	Ala	Tyr	Gln	Ala 600	Thr	Val	Cys	Ala	Arg 605	Ala	Gln	Ala	
CCA	CCT	CCA	TCA	TGG	GAT	CAA	ATG	TGG	AAG	TGT	CTC	ATA	CGG	CTG	AAA	1872
Pro	Pro 610	Pro	Ser	Trp	Asp	Gln 615	Met	Trp	Lys	Cys	Leu 620	Ile	Arg	Leu	Lys	
CCT	ACG	CTG	CAC	GGG	CCA	ACA	CCC	TTG	CTG	TAC	AGG	CTG	GGA	GCC	GTC	1920
Pro 625	Thr	Leu	His	Gly	Pro 630	Thr	Pro	Leu	Leu	Tyr 635	Arg	Leu	Gly	Ala	Val 640	
CAA	AAT	GAG	GTC	ACC	CTC	ACC	CAC	CCC	ATA	ACC	AAA	TAC	ATC .	ATG	GCA	1968
Gln	Asn	Glu	Val	Thr 645	Leu	Thr	His	Pro	Ile 650	Thr	Lys	Tyr	Ile	Met 655	Ala	
	ATG								ACT							1998
Cys	Met	Ser	Ala 660	Asp	Leu	Glu	Val									
			000					665								

(2) INFORMATION FOR SEQ ID NO:110:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 2016 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..2013

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:110:

ATG Met	GGC Gly	AGC Ser	AGC Ser	CAT His	His	CAT His	CAT	CAT	CAC His	Ser	AGC Ser	GGC Gly	CTG Leu	ı Val	CCG Pro		48
CGC Arg	GGC Gly	AGC Ser	CAT His 20	Met	GCT Ala	TAC	TCT Ser	CTG Leu 25	Thr	ACG Thr	GGT Gly	TCT Ser	GTT Val	. Val	ATT Ile		96
GTT Val	GGT Gly	AGA Arg 35	Ile	ATT Ile	TTA Leu	TCT	GGT Gly 40	Ser	GGT Gly	AGT	ATC Ile	ACG Thr 45	Ala	TAC Tyr	TCC	. 1	44
CAA G1n	CAG Gln 50	ACG Thr	CGG Arg	GGC Gly	CTA Leu	CTT Leu 55	GGT Gly	TGC Cys	ATC Ile	ATC Ile	ACT Thr	Ser	CTT	ACA Thr	GGC Gly	1	92
CGG Arg 65	GAC Asp	AAG Lys	AAC Asn	CAG Gln	GTC Val 70	GAG Glu	GGA Gly	GAG Glu	GTT Val	CAG Gln 75	GTG Val	GTT Val	TCC Ser	ACC Thr	GCA Ala 80	2	40
Thr	CAA Gln	Ser	Phe	Leu 85	Ala	Thr	Cys	Val	Asn 90	Gly	Val	Cys	Trp	Thr 95	Val	2	88
Tyr	CAT His	Gly	Ala 100	Gly	Ser	Lys	Thr	Leu 105	Ala	Gly	Pro	Lys	Gly 110	Pro	Ile	3	36
Thr	CAG Gln	Met 115	Tyr	Thr	Asn	Val	Asp 120	Gln	Asp	Leu	Val	Gly 125	Trp	Gln	Ala	3:	84
Pro	Pro 130	Gly	Ala	Arg	Ser	Leu 135	Thr	Pro	Cys	Thr	Cys 140	Gly	Ser	Ser	Asp	41	32
Leu 145	TAC Tyr	Leu	Val	Thr	Arg 150	His	Ala	Asp	Val	Ile 155	Pro	Va1	Arg	Arg	Arg 160	48	30
Gly	GAC Asp	Ser	Arg	Gly 165	Ser	Leu	Leu	Ser	Pro 170	Arg	Pro	Va1	Ser	Tyr 175	Leu	52	88
Lys	GGC Gly	Ser	Ser 180	Gly	Gly	Pro	Leu	Leu 185	Cys	Pro	Ser	Gly	His 190	Ala	Va1	57	6
Gly	ATC Ile	Phe 195	Arg	Ala	Ala	Val	Cys 200	Thr	Arg	Gly	Val	Ala 205	Lys	Ala	Val	62	4
GAC Asp	TTT Phe 210	GTG Val	CCC Pro	GTA Val	GAG Glu	TCC Ser 215	ATG Met	GAA Glu	ACT . Thr	ACT Thr	ATG Met 220	CGG Arg	TCT Ser	CCG Pro	GTC Val	67	'2

	ACG Thr															720
	CAC His															768
	GCA Ala															816
	GCC Ala															. 864
	GAC Asp 290											Thr				912
	GTC Val															960
	GGG Gly															1008
	TCG Ser															1056
	GCT Ala															1104
	GTC Val 370															1152
	GGA Gly															1200
	GGG Gly															1248
	CTC Leu															1296
	CGG Arg															1344
GTC	GTG	GCA	ACA	GAC	GCT	CTG	ATG	ACG	GGC	TAT	ACG	GGC	GAC	TTT	GAC	1392

Val	Val 450	Ala	Thr	Asp	Ala	Leu 455	Met	Thr	Gly	Tyr	Thr 460	Gly	Asp	Phe	Asp	
TCA Ser 465	GTG Val	ATC Ile	GAC Asp	TGT Cys	AAC Asn 470	ACA Thr	TGT Cys	GTC Val	ACC Thr	CAG Gln 475	Thr	GTC Val	GAC Asp	TTC Phe	AGC Ser 480	1440
TTG Leu	GAT Asp	CCC	ACC Thr	TTC Phe 485	ACC Thr	ATT Ile	GAG Glu	ACG Thr	ACG Thr 490	Thr	GTG Val	CCT Pro	CAA Gln	GAC Asp 495	Ala	1488
GTG Val	TCG Ser	CGC Arg	TCG Ser 500	CAG Gln	CGG Arg	CGG Arg	GGT Gly	AGG Arg 505	ACT Thr	GGC Gly	AGG Arg	GGT Gly	AGG Arg 510	AGA Arg	GGC Gly	1536
ATC Ile	TAC Tyr	AGG Arg 515	TTT Phe	GTG Val	ACT Thr	CCG Pro	GGA Gly 520	GAA Glu	CGG Arg	CCC Pro	TCG Ser	GGC Gly 525	ATG Met	TTC Phe	GAT Asp	1584
TCC Ser	TCG Ser 530	GTC Val	CTG Leu	TGT Cys	GAG Glu	TGC Cys 535	TAT Tyr	GAC Asp	GCG Ala	GGC Gly	TGT Cys 540	GCT Ala	TGG Trp	TAC Tyr	GAG Glu	1632
CTC Leu 545	ACC Thr	CCC Pro	GCC Ala	GAG Glu	ACC Thr 550	TCG Ser	GTT Val	AGG Arg	TTG Leu	CGG Arg 555	GCC Ala	TAC Tyr	CTG Leu	AAC Asn	ACA Thr 560	1680
CCA Pro	GGG Gly	TTG Leu	CCC Pro	GTT Val 565	TGC Cys	CAG Gln	GAC Asp	CAC His	CTG Leu 570	GAG Glu	TTC Phe	TGG Trp	GAG Glu	AGT Ser 575	GTC Val	1728
TTC Phe	ACA Thr	GGC Gly	CTC Leu 580	ACC Thr	CAT His	ATA Ile	GAT Asp	GCA Ala 585	CAC His	TTC Phe	TTG Leu	TCC Ser	CAG Gln 590	ACC Thr	AAG Lys	1776
CAG Gln	GCA Ala	GGA Gly 595	GAC Asp	AAC Asn	TTC Phe	CCC Pro	TAC Tyr 600	CTG Leu	GTA Val	GCA Ala	TAC Tyr	CAA Gln 605	GCC Ala	ACG Thr	GTG Val	1824
TGC Cys	GCC Ala 610	AGG Arg	GCT Ala	CAG Gln	GCC Ala	CCA Pro 615	CCT Pro	CCA Pro	TCA Ser	TGG Trp	GAT Asp 620	CAA Gln	ATG Met	TGG Trp	AAG Lys	1872
TGT Cys 625	CTC Leu	ATA Ile	CGG Arg	CTG Leu	AAA Lys 630	CCT Pro	ACG Thr	CTG Leu	CAC His	GGG Gly 635	CCA Pro	ACA Thr	CCC Pro	TTG Leu	CTG Leu 640	1920
TAC Tyr	AGG Arg	CTG Leu	GGA Gly	GCC Ala 645	GTC Val	CAA Gln	AAT Asn	GAG Glu	GTC Val 650	ACC Thr	CTC Leu	ACC Thr	CAC His	CCC Pro 655	ATA Ile	1968
ACC Thr	AAA Lys	TAC Tyr	ATC Ile 660	ATG Met	GCA Ala	TGC Cys	ATG Met	TCG Ser 665	GCT Ala	GAC Asp	CTG Leu	GAG Glu	GTC Val 670	GTC Val	•	2013
ACT																2016

(2) INFORMATION FOR SEQ ID NO:111:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 2016 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

(A) NAME/KEY: CDS (B) LOCATION: 1..2013

(xi) SE	QUEN	CE D	ESCR	IPTI	ON:	SEQ	ID N	ю:11	1:			
GGC Gly												Pro	48
GGC Gly								Thr					96
GGT Gly													144
CAG Gln 50													192
GAC Asp													240
CAA Gln													288
CAT His													336
CAG Gln													384
CCC Pro 130													432

155

480

160

150

145

CTT TAC TTG GTC ACG AGA CAT GCT GAC GTC ATT CCG GTG CGC CGG CGG

Leu Tyr Leu Val Thr Arg His Ala Asp Val Ile Pro Val Arg Arg Arg

	GAC Asp									Arg					Leu		528
AA(Lys	GGC Gly	TCT	GCT Ala 180	GGT Gly	GGT Gly	CCA Pro	CTG Leu	CTC Leu 185	Cys	CCT Pro	TCG Ser	GGG Gly	CAC His 190	Ala	GTG Val	r	576
	ATC							Thr					Lys				624
GA0 Ası	Phe 210	GTG Val	CCC Pro	GTA Val	GAG Glu	TCC Ser 215	ATG Met	GAA Glu	ACT Thr	ACT Thr	ATG Met 220	Arg	TCT Ser	CCG Pro	GTC Val	•	672
	ACG Thr																720
	CAC His																768
	GCA Ala																816
	GCC Ala																864
	GAC Asp 290																912
	GTC Val																960
TCT Ser	GGG Gly	GGC Gly	GCT Ala	TAT Tyr 325	GAC Asp	ATC Ile	ATA Ile	ATA Ile	TGT Cys 330	GAT Asp	GAG Glu	TGC Cys	CAT His	TCA Ser 335	ACT Thr		1008
	TCG Ser																1056
	GCT Ala																1104
TCG Ser	GTC Val 370	ACC Thr	GTG Val	CCA Pro	CAC His	CCA Pro 375	AAC Asn	ATC Ile	GAG Glu	GAG Glu	GTG Val 380	GCC Ala	CTG Leu	TCT Ser	AAT Asn		1152
ACT	GGA	GAG	ATC	ccc	TTC	TAT	GGC	AAA	GCC	ATC	ccc	ATT	GAA	GCC	ATC		1200

Thr 385	Gly	Glu	Ile	Pro	Phe 390	Tyr	Gly	Lys	Ala	Ile 395	Pro	Ile	Glu	Ala	Ile 400	
					CTC Leu											1248
					CTG Leu											1296
					GTG Val										GTT Val	1344
					GCT Ala											1392
					AAC Asn 470											1440
					ACC Thr											1488
					CGG Arg											1536
					ACT Thr											1584
					GAG Glu							Ala				1632
					ACC Thr 550											1680
					TGC Cys											1728
					CAT His											1776
					TTC Phe										GTG Val	1824
					GCC Ala							Gln				1872

	CTC Leu														1920
	AGG Arg														1968
	AAA Lys														2013
ACT															2016
(2)	(ii) SE ((((()) MO	QUEN A) L B) T C) S D) T LECU ATUF A) N	CE C ENGT YPE: TRAN OPOL	HARA H: 6 nuc DEDN OGY: YPE:	CTER 48 b 1eic ESS: 1in cDN	ISTI ase aci sin ear	CS: pair d	s						
	(xi) SE	QUEN	CE D	ESCR	IPTI	ON:	SEQ	ID N	0:11	2:				
	GGC Gly														48
	GGC Gly														96
	CCT Pro							Ser							144
	TGC Cys 50						Thr					Asn		GAG Glu	192
	GAG Glu														240
	GTC Val				Cys					His				Lys	288

	TTA Leu									336
	CAG Gln									384
	CCA Pro 130									432
	GAC Asp								•	480
	TCC Ser									528
	CTC Leu									576
	ACC Thr									624
	GAA Glu 210			*						648

- (2) INFORMATION FOR SEQ ID NO:113:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 648 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: cDNA
 - (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION: 1..640
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:113

ATG GGC AGC AGC CAT CAT CAT CAT CAC AGC AGC GGC CTG GTG CCG 48 Met Gly Ser Ser His His His His His Ser Ser Gly Leu Val Pro 1 5 15 96

CGC GGC AGC CAT ATG GGT TCT GTT GTT ATT GTT GGT AGA ATT ATT TTA

	Arg	Gly	Ser	His 20	Met	Gly	Ser	Val	Val 25	Ile	Val	Gly	Arg	Ile 30	Ile	Leu	
						ACG Thr											144
						AGC Ser							Asn				192
						GTT Val 70											240
						TGT Cys											288
						AAG Lys											336
						GGC Gly											384
						GGC Gly											432
i						GTG Val 150											480
						GTC Val											528
						GGG Gly											576
						GCG Ala											624
		GAA Glu	Thr			c e	GTCI	TGA									648

(2) INFORMATION FOR SEQ ID NO:114:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 498 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 1..498

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:114:

	ATC Ile															. 48
1				5					10					15		
	CAG Gln															96
			20					25		,			30			
	GGC Gly															144
	_	35	_	_			40		_		_	45	_			
	GGC Gly															192
	50					55					60					
	CTC Leu															240
65	200	V.4.1	CII	LLD	70	1114	110	110	OLY	75	111.9	DCI	Dea	****	80	
	ACC Thr															288
Cys	1111	Cys	GIY	85	ser	мыр	Leu	TĂT	90	Val	THE	Arg	nıs	95	ASP	
	ATT Ile															336
Vai	116	FIO	100	Arg	nrg	nrg	Giy	105	261	AIG	GIŞ	361	110	Бец	Sei	
	AGG Arg															384
110	1119	115	va1	DCI	-7-	Бей	120	OLY	DCI	DOL	GLY	125	110	Deu	Deu	
	CCT Pro															432
Cys	130	Ser	GIY	1113	ліа	135	GLY	116	rne	AIG	140	лта	vai	Cys	1111	
	GGG Gly															480
145	GIY	741	.114	د ړ د	150	val	лор	2116	vai	155	vai	JIU	201	net	160	
	ACT Thr				TGA											498
			9	165												

(2) INFORMATION FOR SEQ ID NO:115:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 648 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION: 1..648

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:115:

GGC Gly								48
GGC Gly								96
GGT Gly								144
TGC Cys 50								192
GAG Glu								240
GTC Val								288
TTA Leu								336
CAG Gln								384
CCA Pro 130								432
GAC Asp								480

CTC Leu	TCC Ser	CCC	AGG Arg	CCT Pro 165	GTC Val	TCC Ser	TAC Tyr	TTG Leu	AAG Lys 170	GGC Gly	TCT Ser	TCG Ser	GGT Gly	GGT Gly 175	CCA Pro	5	528
	CTC Leu															. !	576
	ACC Thr															•	624
	GAA Glu 210						*									. 6	648
(2)	INFO	ORMAT	NOI	FOR	SEQ	ID N	0:11	6:									
	(i	(QUEN A) L B) T C) S D) T	ENGT YPE: TRAN	H: 2 nuc DEDN	007 leic ESS:	base aci sin	pai d	rs								
	(ii) MO	LECU	LE T	YPE:	CDN	A										
	(ix		EATUF A) N B) L	AME/													
	(xi) SE	QUEN	CE D	ESCR:	IPTI	ON:	SEQ	ID N	0:11	6:						
	CAT His																48
	ATC Ile																96
	ACT Thr															1	144
Ile CAG		Ser 35 GTT	Leu	Thr	Gly GCA	Arg ACA	Asp 40 CAA	Lys TCC	Asn TTC	Gln CTG	Val GCG	Glu 45 ACC	Gly TGC	Glu GTC	Val AAC		144 192

GGC CCA AAG GGG CCA ATC ACC CAG ATG TAC ACT AAT GTG GAC CAG GAC 288

Gly	Pro	Lys	Gly	Pro 85	Ile	Thr	Gln	Met	Tyr 90	Thr	Asn	Val	Asp	Gln 95	Asp	
	GTC Val															336
	TGT Cys															384
	CCG Pro 130														ccc Pro	432
	CCT Pro															480
	TCG Ser															528
	GTT Val															576
	ATG Met															624
	CAG Gln 210															672
	AGT Ser															720
	GTC Val															768
	TCT Ser															816
	ATT Ile															864
	GCC Ala 290															912
	GAG Glu															960

		CTG Leu															1008
		GCT Ala															1056
		GTG Val															1104
		CCC Pro 370											Ile			CAT His	1152
		AAG Lys															1200
		AAC Asn															1248
		ATC Ile															1296
		ACG Thr															1344
		ACA Thr 450															1392
		GTG Val															1440
		AGG Arg															1488
		TCG Ser															1536
		TGT Cys															1584
		GCC Ala 530															1632
(GAG	TTC	TGG	GAG	AGT	GTC	TTC	ACA	GGC	CTC	ACC	CAT	ATA	GAT	GCA	CAC	1680

Glu 545	Phe	Trp	Glu	Ser	Va1 550	Phe	Thr	Gly	Leu	Thr 555	His	Ile	Asp	Ala	His 560	
	TTG Leu															1728
	TAC Tyr															1776
	GAT Asp															1824
	CCA Pro 610															1872
	CTC Leu															1920
	CTG Leu															1968
	CGC Arg								Pro							2007

- (2) INFORMATION FOR SEQ ID NO:117:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 2007 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: cDNA
 - (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION: 1..2004
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:117:

ATG	CAT	ATG	CAT	CAT	CAT	CAC	CAT	CAT	CTG	GIG	CCG	CGC	GGC	AGC	GCG	48
Met	His	Met	His	His	His	His	His	His	Leu	Val	Pro	Arg	Gly	Ser	Ala	
1				5					10					15		
ccc	a mc	3.00	000	mac.	maa	C2.2	CAC	200	ccc	ccc	CIDA	Cmm	CCM	maa	A TO C	96

Pro Ile Thr Ala Tyr Ser Gln Gln Thr Arg Gly Leu Leu Gly Cys Ile

25 AAG ACT AGC CTT ACA GGC CGG GAC AAG AAC CAG GTC GAG GGA GAG GTT

20

Lys Thr Ser Leu Thr Gly Arg Asp Lys Asn Gln Val Glu Gly Glu Val 40 CAG GTG GTT TCC ACC GCA ACA CAA TCC TTC CTG GCG ACC TGC GTC AAC 192 Gln Val Val Ser Thr Ala Thr Gln Ser Phe Leu Ala Thr Cys Val Asn GGC GTG TGT TGG ACC GTT TAC CAT GGT GCT GGC TCA AAG ACC TTA GCC Gly Val Cys Trp Thr Val Tyr His Gly Ala Gly Ser Lys Thr Leu Ala 65 7.0 75 GGC CCA AAG GGG CCA ATC ACC CAG ATG TAC ACT AAT GTG GAC CAG GAC 288 Gly Pro Lys Gly Pro Ile Thr Gln Met Tyr Thr Asn Val Asp Gln Asp 90 CTC GTC GGC TGG CAG GCG CCC CCC GGG GCG CGT TCC TTG ACA CCA TGG 336 Leu Val Gly Trp Gln Ala Pro Pro Gly Ala Arg Ser Leu Thr Pro Cys 100 105 ACC TGT GGC AGC TCA GAC CTT TAC TTG GTC ACG AGA CAT GCT GAC GTC 384 Thr Cys Gly Ser Ser Asp Leu Tyr Leu Val Thr Arg His Ala Asp Val 120 115 ATT CCG GTG CGC CGG CGC GAC AGT AGG GGG AGC CTG CTC CCC 432 Ile Pro Val Arg Arg Gly Asp Ser Arg Gly Ser Leu Leu Ser Pro 130 135 140 AGG CCT GTC TCC TAC TTG AAG GGC TCT TCG GGT GGT CCA CTG CTC TGC 480 Arg Pro Val Ser Tyr Leu Lys Gly Ser Ser Gly Gly Pro Leu Leu Cys 145 150 155 CCT TCG GGG CAC GCT GTG GGC ATC TTC CGG GCT GCC GTA TGC ACC CGG Pro Ser Gly His Ala Val Gly Ile Phe Arg Ala Ala Val Cys Thr Arg 165 170 GGG GTT GCG AAG GCG GTG GAC TTT GTG CCC GTA GAG TCC ATG GAA ACT 576 Gly Val Ala Lys Ala Val Asp Phe Val Pro Val Glu Ser Met Glu Thr 180 ACT ATG CGG TCT CCG GTC TTC ACG GAC AAC TCA TCC CCC CCG GCC GTA Thr Met Arg Ser Pro Val Phe Thr Asp Asn Ser Ser Pro Pro Ala Val 195 200 CCG CAG TCA TTT CAA GTG GCC CAC CTA CAC GCT CCC ACT GGC AGC GGC 672 Pro Gln Ser Phe Gln Val Ala His Leu His Ala Pro Thr Gly Ser Gly 210 215 AAG AGT ACT AAA GTG CCG GCT GCA TAT GCA GCC CAA GGG TAC AAG GTG 720 Lys Ser Thr Lys Val Pro Ala Ala Tyr Ala Ala Gln Gly Tyr Lys Val 230 235

CTC GTC CTC AAT CCG TCC GTT GCC GCT ACC TTA GGG TTT GGG GCG TAT Leu Val Leu Asn Pro Ser Val Ala Ala Thr Leu Gly Phe Gly Ala Tyr 245 250

A	TG et	TCT Ser	AAG Lys	GCA Ala 260	CAC His	GGT Gly	ATT Ile	GAC Asp	CCC Pro 265	Asn	ATC Ile	AGA Arg	ACT Thr	GGG Gly 270	GTA Val	AGG Arg	816
						GCC Ala										TTT Phe	864
						TGC Cys							Ile				912
A						ACT Thr 310										ACA Thr 320	. 960
						GAG Glu											1008
						GGA Gly											1056
						AAT Asn											1104
						ATC Ile											1152
S						GAC Asp 390											1200
						TAT Tyr											1248
						GTT Val											1296
						GAC Asp											1344
						AGC Ser											1392
T						GCA Ala 470											1440
						GGC Gly											1488

485 490 495 CCC TCG GGC ATG TTC GAT TCC TCG GTC CTG TGT GAG TGC TAT GAC GCG Pro Ser Gly Met Phe Asp Ser Ser Val Leu Cys Glu Cys Tyr Asp Ala 505 GGC TGT GCT TGG TAC GAG CTC ACC CCC GCC GAG ACC TCG GTT AGG TTG 1584 Gly Cys Ala Trp Tyr Glu Leu Thr Pro Ala Glu Thr Ser Val Arg Leu 515 520 CGG GCC TAC CTG AAC ACA CCA GGG TTG CCC GTT TGC CAG GAC CAC CTG Arg Ala Tyr Leu Asn Thr Pro Gly Leu Pro Val Cys Gln Asp His Leu 530 535 540 GAG TTC TGG GAG AGT GTC TTC ACA GGC CTC ACC CAT ATA GAT GCA CAC 1680 Glu Phe Trp Glu Ser Val Phe Thr Gly Leu Thr His Ile Asp Ala His 545 550 TTC TTG TCC CAG ACC AAG CAG GCA GGA GAC AAC TTC CCC TAC CTG GTA 1728 Phe Leu Ser Gln Thr Lys Gln Ala Gly Asp Asn Phe Pro Tyr Leu Val 565 570 GCA TAC CAA GCC ACG GTG TGC GCC AGG GCT CAG GCC CCA CCT CCA TCA 1776 Ala Tyr Gln Ala Thr Val Cys Ala Arg Ala Gln Ala Pro Pro Pro Ser 585 TGG GAT CAA ATG TGG AAG TGT CTC ATA CGG CTG AAA CCT ACG CTG CAC 1824 Trp Asp Gln Met Trp Lys Cys Leu Ile Arg Leu Lys Pro Thr Leu His 595 600 605 GGG CCA ACA CCC TTG CTG TAC AGG CTG GGA GCC GTC CAA AAT GAG GTC 1872 Gly Pro Thr Pro Leu Leu Tyr Arg Leu Gly Ala Val Gln Asn Glu Val 610 615 1920 Thr Leu Thr His Pro Ile Thr Lys Tyr Ile Met Ala Cys Met Ser Ala 625 630 635 640 GAC CTG GAG GTC GTT ACG TAG GAA TTC GAG CTC CGT CGA CAA GCT TGC 1968 Asp Leu Glu Val Val Thr * Glu Phe Glu Leu Arg Arg Gln Ala Cys 645 650 GGC CGC ACT CGA GCA CCA CCA CCA CCA CTG AGA TCC 2007 Gly Arg Thr Arg Ala Pro Pro Pro Pro Pro Leu Arg 660 665

- (2) INFORMATION FOR SEQ ID NO:118:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 2007 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: cDNA

- (ix) FEATURE:
- (A) NAME/KEY: CDS
- (B) LOCATION: 1..2004

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:118:

(11	, 55	COPM	CE D	BOCK	TEIT	OIV.	SEQ	יו עד	0.11	٠.			
CAT His													48
ATC Ile													96
ACT Thr													144
GTG Val 50													192
GTG Val													240
CCA Pro													288
GTC Val													336
TGT Cys													384
CCG Pro 130													432
CCT Pro													480
TCG Ser													528
GTT Val													576
ATG Met													624

	CAG Gln 210															672
	AGT Ser															720
	GTC Val															768
	TCT Ser															816
	ATT Ile															864
	GCC Ala 290															912
	GAG Glu															960
	CTG Leu															1008
	GCT Ala															1056
	GTG Val															1104
	CCC Pro 370															1152
	AAG Lys															1200
	AAC Asn															1248
	ATC Ile															1296
TAT	ACG	GGC	GAC	TTT	GAC	TCA	GTG	ATC	GAC	TGT	AAC	ACA	TGT	GTC	ACC	1344

Tyr	Thr	Gly 435	Asp	Phe	Asp	Ser	Val 440	Ile	Asp	Cys	Asn	Thr 445	Cys	Val	Thr	
	ACA Thr 450											Ile				1392
	GTG Val															1440
	AGG Arg														CGG Arg	1488
	TCG Ser															1536
	TGT Cys															1584
	GCC Ala 530															1632
	TTC Phe															1680
	TTG Leu															1728
	TAC Tyr															1776
	GAT Asp															1824
	CCA Pro 610											Gln				1872
	CTC Leu															1920
	CTG Leu															1968
	CGC Arg								Pro							2007

(2) INFORMATION FOR SEQ ID NO:119:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 2007 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
 - (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION: 1..2004

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:119:

	CAT His															48
	ATC Ile															96
	ACT Thr															144
	GTG Val 50															192
	GTG Val															240
	CCA Pro															288
	GTC Val															336
	TGT Cys															384
	CCG Pro 130															432
AGG	CCT	GTC	TCC	TAC	TTG	AAG	GGC	TCT	TCG	GGT	GGT	CCA	CTG	CTC	TGC	480

Ara	Pro	Val	Ser	Tvr	Leu	Tare	Glv	Ser	Sar	Gly	Glv	Pro	Tou	Lou	Cyro	
145	110	V 4 1	DCI	131	150	шуз	GLY	Ser	Ser	155	GIĀ	FIO	Leu	Leu	160	
															CGG Arg	528
					GTG Val									Glu	ACT Thr	576
					GTC Val											624
					GTG Val											672
					CCG Pro 230											720
					TCC Ser											768
					GGT Gly											816
					GCC Ala											864
					TGC Cys											912
					ACT Thr 310											960
					GAG Glu											1008
					GGA Gly											1056
					AAT Asn											1104
					ATC Ile											1152

															GGA Gly 400		1200
			GTG Val												Pro		1248
ACT Thr	TCC Ser	GGA Gly	GAC Asp 420	GTC Val	GTT Val	GTC Val	GTG Val	GCA Ala 425	ACA Thr	GAC Asp	GCT Ala	CTG Leu	ATG Met 430	Thr	GGC Gly		1296
			Asp												ACC Thr	•	1344
			GAC Asp														1392
			CAA Gln														1440
			AGG Arg														1488
CCC Pro	TCG Ser	GGC Gly	ATG Met 500	TTC Phe	GAT Asp	TCC Ser	TCG Ser	GTC Val 505	CTG Leu	TGT Cys	GAG Glu	TGC Cys	TAT Tyr 510	GAC Asp	GCG Ala		1536
			TGG Trp														1584
			CTG Leu														1632
			GAG Glu														1680
			CAG Gln														1728
			GCC Ala 580														1776
			ATG Met														1824
GGG	CCA	ACA	ccc	TTG	CTG	TAC	AGG	CTG	GGA	GCC	GTC	CAA	AAT	GAG	GTC		1872

Gly	Pro 610	Thr	Pro	Leu	Leu	Tyr 615	Arg	Leu	Gly	Ala	Val 620	Gln	Asn	Glu	Val	
	CTC Leu															1920
	CTG Leu															1968
	CGC Arg											TCC				2007
(2)	INFO) SE (.	QUEN A) L B) T	CE C ENGT YPE:	HARA H: 2	ID N CTER 007 leic ESS:	ISTI base aci	CS: pai đ	rs							

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..2007

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:120:

CAT His								48
ATC Ile								96
ACT Thr								144
GTG Val 50						Thr		192
GTG Val								240
CCA Pro				 	 		 	 288

GTC Val		Gln			Ala				3	336
TGT Cys				Leu			Ala		3	384
CCG Pro 130									4	132
CCT Pro						Gly				180
TCG Ser									Ş	528
GTT Val									5	576
ATG Met									6	524
CAG Gln 210									6	72
AGT Ser									7	20
GTC Val									7	68
TCT Ser									8	16
ATT Ile									8	64
GCC Ala 290									9	12
GAG Glu									9	60
CTG Leu									10	800

325 330 335

ACC Thr	GCT Ala	ACG Thr	CCT Pro 340	CCG Pro	GGA Gly	TCG	GTC Val	ACC Thr 345	GTG Val	CCA Pro	CAC His	CCA Pro	AAC Asn 350	ATC Ile	GAG Glu	1056
	GTG Val							Glu					Gly			1104
	CCC Pro 370											Ile				1152
	AAG Lys															1200
	AAC Asn															1248
ACT Thr	ATC Ile	GGA Gly	GAC Asp 420	GTC Val	GTT Val	GTC Val	GTG Val	GCA Ala 425	ACA Thr	GAC Asp	GCT Ala	CTG Leu	ATG Met 430	ACG Thr	GGC Gly	1296
TAT Tyr	ACG Thr	GGC Gly 435	GAC Asp	TTT Phe	GAC Asp	TCA Ser	GTG Val 440	ATC Ile	GAC Asp	TGT Cys	AAC Asn	ACA Thr 445	TGT Cys	GTC Val	ACC Thr	1344
CAG Gln	ACA Thr 450	GTC Val	GAC Asp	TTC Phe	AGC Ser	TTG Leu 455	GAT Asp	CCC Pro	ACC Thr	TTC Phe	ACC Thr 460	ATT Ile	GAG Glu	ACG Thr	ACG Thr	1392
	GTG Val															1440
	AGG Arg															1488
	TCG Ser															1536
GGC Gly	TGT Cys	GCT Ala 515	TGG Trp	TAC Tyr	GAG Glu	CTC Leu	ACC Thr 520	CCC Pro	GCC Ala	GAG Glu	ACC Thr	TCG Ser 525	GTT Val	AGG Arg	TTG Leu	1584
	GCC Ala 530															1632
GAG Glu 545	TTC Phe	TGG Trp	GAG Glu	AGT Ser	GTC Val 550	TTC Phe	ACA Thr	GGC Gly	CTC Leu	ACC Thr 555	CAT His	ATA Ile	GAT Asp	GCA Ala	CAC His 560	1680

TTG Leu								1728
TAC Tyr								1776
GAT Asp								1824
CCA Pro 610							GTC Val	1872
CTC Leu								1920
CTG Leu								1968
CGC Arg				Pro				2007

- (2) INFORMATION FOR SEQ ID NO:121:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 36 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: RNA (genomic)
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:121:

GCUCGCCCGG GGAUCCUCUA GGAAUACACG UUCGAU 36

- (2) INFORMATION FOR SEQ ID NO:122:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 36 base pairs
 - (B) TYPE: nucleic acid (C) STRANDEDNESS: single
 - (C) STRANDEDNESS: SINGLE
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: RNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:122:

CUAGAGGAUC CCCGGGCGAG CCCUAUAGUG AGUCGU 36

- (2) INFORMATION FOR SEQ ID NO:123:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 21 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single (D) TOPOLOGY; linear

 - (ii) MOLECULE TYPE: cDNA
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:123:

GCTCGCCCGG GGATCCTCTA G

21

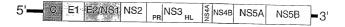
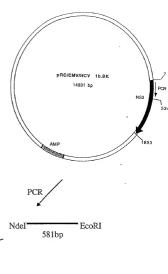


Fig. 1



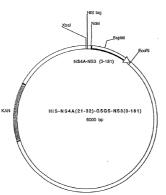
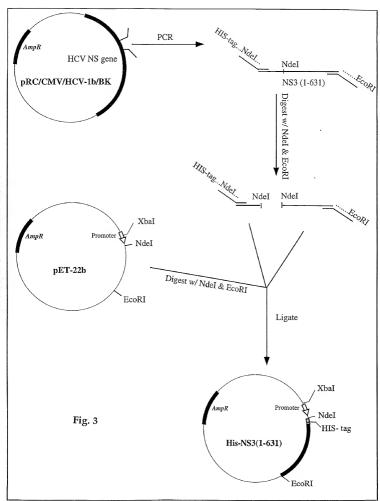
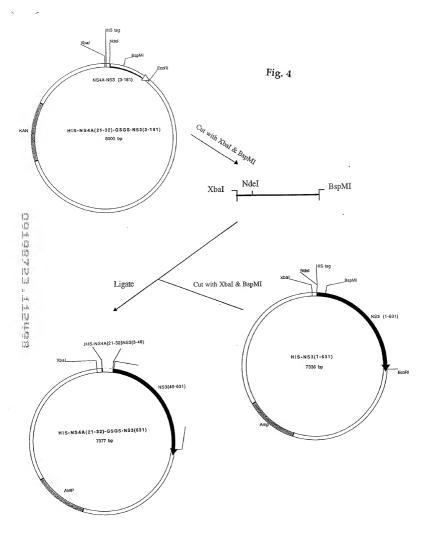
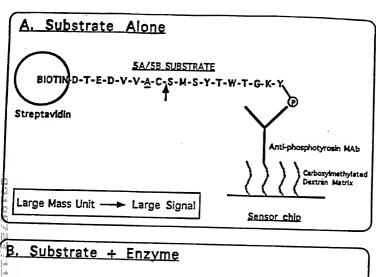


Fig. 2







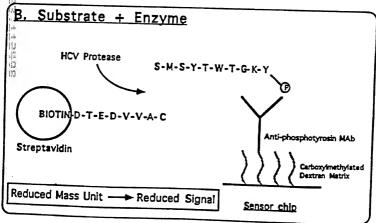


Figure 6

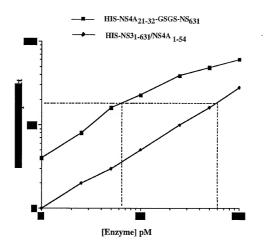
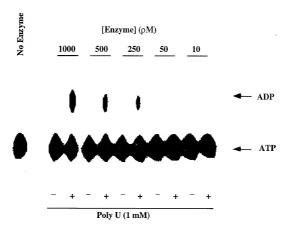


Figure 7



DECLARATION AND POWER OF ATTORNEY FOR PROVISIONAL PATENT APPLICATION

Attorney's Docket No. __JB0800_

As a below-named inventor, I hereby declare that:

My residence, post office address and citizenship are as stated below next to my name;

I believe I am the original, first sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural names are listed below) of the subject matter which is claimed and for which a patent is sought on the invention entitled:

SINGLE-CHAIN RECOMBINANT COMPLEXES OF HEPATITIS C VIRUS NS3 PROTEASE AND NS4A COFACTOR PEPTIDE

the specification of wh	ich is attached hereto.		
specification, including	ave reviewed and understage the claims.	and the contents of the	above-identified
(i) A acknowledge the duty in accordance in	y to disclose information v nce with 37 C.F.R. §1.56(which is material to the pa).	atentability of this
I hereby claim foreign any foreign application defined below any foreign	priority benefits under Titl n(s) for patent or inventor! oreign application for pate application on which prio	s certificate listed below nt or inventor's certificat	and have also
Prior Foreign Application(s):			Priority Claimed
(Number)	(Country)	(Day/Month/Year Filed)	Yes or No
refeby claim the ben provisional application 60/067,315 (Application Number)	efit under Title 35, United (s) listed below: November 28, 1997 (Filing Date)	States Code, §119(e) c	f any United States
(Application Number)	July 28, 1998 (Filing Date)		
application(s) listed be application is not discluding by the first paragraph disclose material inform	nefit under Title 35, Unite elow and, insofar as the solosed in the prior United of Title 35, United State mation as defined in Title en the filling date of the e of this application:	subject matter of each of States application in the s Code, §112, I ackno 37, Code of Federal Re	of the claims of this e manner provided wledge the duty to equiations, §1.56(a)
(Application Serial No.)	(Filing Date)	(Status – patented, pend	ing, abandoned)

Power of Attorney: As a named inventor, I hereby appoint the following attorney(s) and/or agent(s) to prosecute this application and transact all business in the Patent and

agorit(c) to proceed the application ar	ia nanoaot an basiness in the ratein
Trademark Office connected therewith.	(List name and registration number.)

Carl W. Battle	Reg. No. 30731	John J. Maitner	Reg. No. 25636
Edwin P. Ching	Reg. No. 34090	Joseph T. Majka	Reg. No. 30570
Eric S. Dicker	Reg. No. 31699	Arthur Mann	Reg. No. 35598
Norman C. Dulak	Reg. No. 31608	Edward H. Mazer	Reg. No. 27573
Cynthia L. Foulke	Reg. No. 32364	Jave P. McLaughlin	Reg. No. 41211
Robert A. Franks	Reg. No. 28605	Sheela Mohan-Peterson	Reg. No. 41201
James M. Gould	Reg. No. 33702	Richard B. Murphy	Reg. No. 35296
Richard J. Grochala	Reg. No. 31518	James R. Nelson	Reg. No. 27929
Thomas D. Hoffman	Reg. No. 28221	Immac J. Thampoe	Reg. No. 36322
Henry C. Jeanette	Reg. No. 30856	Paul A. Thompson	Reg. No. 35385
Susan Lee	Reg. No. 30653	Donald W. Wyatt	Reg. No. 40,876
Anita W Magatti	Reg No. 29825	•	•

Correspondence	

Jaye P. McLaughlin Schering-Plough Corporation Patent Department, K-6-1, 1990 2000 Galloping Hill Road Direct Telephone Calls to:

Name: Jaye P. McLaughlin Telephone No.: (908) 298-5056 Facsimile No.: (908) 298-5388

Kenilworth, New Jersey 07033-0530

FULL NAME OF 1ST	FAMILY NAME	FIRST GIVEN NAME	SECOND GIVEN NAME
INVENTOR	Malcolm	Bruce	A.
RESIDENCE & CITIZENSHIP	CITY	STATE OR FOREIGN COUNTRY	COUNTRY OF CITIZENSHIP
11	Westfield	New Jersey	USA
*POST OFFICE ADDRESS	POST OFFICE ADDRESS	CITY	STATE & ZIP CODE/COUNTRY
19	515 Trinity Place	Westfield	New Jersey 07090 USA

18	net.			
P liping.	FULL NAME OF 2ND JOINT INVENTOR	FAMILY NAME	FIRST GIVEN NAME	SECOND GIVEN NAME
14	•	Taremi	S. Shane	
	RESIDENCE & CITIZENSHIP	CITY	STATE OR FOREIGN COUNTRY	COUNTRY OF CITIZENSHIP
		Upper Montclair	New Jersey	Iran
	POST OFFICE ADDRESS	POST OFFICE ADDRESS	CITY	STATE & ZIP CODE/COUNTRY
	l .	12 Park Terrace	Upper Montclair	New Jersey 07043 LISA

FULL NAME OF 3RD	FAMILY NAME	FIRST GIVEN NAME	SECOND GIVEN NAME
JOINT INVENTOR			
1	Weber	Patricia	C.
RESIDENCE &	CITY	STATE OR FOREIGN COUNTRY	COUNTRY OF CITIZENSHIP
CITIZENSHIP			
	Yardley	Pennsylvania	USA
POST OFFICE	POST OFFICE ADDRESS	CITY	STATE & ZIP CODE/COUNTRY
ADDRESS			
	1970 Timber Lakes Drive	Yardlev	Pennsylvania 19067 USA

FULL NAME OF 4TH JOINT INVENTOR	FAMILY NAME	FIRST GIVEN NAME	SECOND GIVEN NAME
	Yao	Nanhua	
RESIDENCE & CITIZENSHIP	CITY	STATE OR FOREIGN COUNTRY	COUNTRY OF CITIZENSHIP
	Edison	New Jersey	Peoples Republic of China
POST OFFICE ADDRESS	POST OFFICE ADDRESS	CITY	STATE & ZIP CODE/COUNTRY
	4 Timothy Court	Edison	New Jersey 08837 USA

I further declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code, and that such willful false statements may jeopardize the validity of the application or any patent issuing thereon.

Signature of First Inventor	Signature of Second Inventor	Signature of Third Inventor
Duce Malrotin	_S. Shane Care.	- Pahvaia C. Wiles
Date XI = 18 - 98	Date 11 - 18 - 98	Date 11-18-98
Bruce A. Malcolm	S. Shane Taremi	Patricia C. Weber
Signature of Fourth Inventor		

Nanhua Yao

Date

Rev. 1/96 JHB